

# Density Tracking by Quadrature for Stochastic Differential Equations

Harish S. Bhat<sup>\*†</sup>

R. W. M. A. Madushani<sup>\*</sup>

December 19, 2016

## Abstract

We develop and analyze a method, density tracking by quadrature (DTQ), to compute the probability density function of the solution of a stochastic differential equation. The derivation of the method begins with the discretization in time of the stochastic differential equation, resulting in a discrete-time Markov chain with continuous state space. At each time step, the DTQ method applies quadrature to solve the Chapman-Kolmogorov equation for this Markov chain. In this paper, we focus on a particular case of the DTQ method that arises from applying the Euler-Maruyama method in time and the trapezoidal quadrature rule in space. Our main result establishes that the density computed by the DTQ method converges in  $L^1$  to both the exact density of the Markov chain (with exponential convergence rate), and to the exact density of the stochastic differential equation (with first-order convergence rate). We also establish a Chernoff bound that implies convergence of a domain-truncated version of the DTQ method. We carry out numerical tests to show that the empirical performance of the DTQ method matches theoretical results, and also to demonstrate that the DTQ method can compute densities several times faster than a Fokker-Planck solver, for the same level of error.

<sup>\*</sup>Applied Mathematics Unit, University of California, Merced, CA 95343

<sup>†</sup>email: hbhat@ucmerced.edu

# 1 Introduction

Consider the stochastic differential equation (SDE) for the scalar process  $X_t$ ,

$$dX_t = f(X_t)dt + g(X_t)dW_t, \quad (1)$$

where  $W_t$  is the Wiener process.  $X_t$  is an Itô diffusion; neither the drift  $f$  nor the diffusion  $g$  feature explicit time-dependence. Assuming regularity of  $f$  and  $g$ , the process  $X_t$  has a probability density function  $p(x, t)$  [27]. In this paper, we develop a convergent numerical method to solve for  $p$ . For ease of reference, we call our method DTQ (density tracking by quadrature).

To introduce the DTQ method informally, let us describe the three main steps in its derivation:

1. Discretize the SDE (1) in time using a convergent stochastic time-stepping method.
2. Interpret the time-discretized equation as a discrete-time Markov chain on a continuous state space; let  $\tilde{p}$  denote its probability density function. We can then write down a Chapman-Kolmogorov equation that enables us to evolve  $\tilde{p}$  forward in time.
3. Discretize both the Chapman-Kolmogorov equation and  $\tilde{p}$  in space, e.g., using a spatial grid and numerical quadrature. Let  $\hat{p}$  denote the discrete-space approximation of  $\tilde{p}$ .

We emphasize that these steps form a framework that encompasses many possible algorithms. In this paper, we use the explicit Euler-Maruyama method in step 1 and the trapezoidal rule in step 3; unless stated otherwise, this is the DTQ method analyzed here. Had we made different choices in these steps, we would have obtained a different method in the DTQ family.

In this paper, we prove that  $\hat{p}$  converges to  $p$  as the discretization parameters tend to zero. Because there are existing results on the convergence of  $\tilde{p}$  to  $p$ , the main task of this paper is to show that  $\hat{p} \rightarrow \tilde{p}$ .

More specifically, the foundational work of Bally and Talay [2] established conditions under which  $\tilde{p}$  converges to  $p$ , in the case where the Euler-Maruyama method is used to discretize the SDE (1) in time. Let  $\|f\|_1$  denote the  $L^1$  norm of a function  $f$ . Suppose we seek the density of (1) at time  $T > 0$ . Let  $h > 0$  denote the temporal step size; as we take  $h \rightarrow 0$ , we assume  $T = Nh$  stays fixed. Then the results of [2] imply that  $\|p(\cdot, T) - \tilde{p}(\cdot, T)\|_1 = O(h)$ .

Our work builds on this result. The DTQ method analyzed here combines Euler-Maruyama temporal discretization with the trapezoidal rule on an equispaced grid. This results in a fast, simple method to compute an approximation  $\hat{p}$  such that  $\|\tilde{p}(\cdot, T) - \hat{p}(\cdot, T)\|_1 = O(h^{-1} \exp(-rh^{-\kappa}))$  for positive constants  $r, \kappa$ . The user of the method can control  $\kappa$  by adjusting the relationship between the spatial and temporal grid spacings.

The primary application of this work that we envision is in statistical inference for diffusion processes. The DTQ method can be used to numerically approximate the likelihood function for a diffusion that is observed at discrete points in time. In fact, we have already begun to use the DTQ method to devise both Bayesian and frequentist inference algorithms [7, 8]. The present work lays a theoretical foundation for these statistical applications. Additionally, note that when inference procedures for diffusions have been compared, a method that approximates the likelihood by numerically solving the Fokker-Planck (or Kolmogorov) equation achieves superior accuracy at the cost of excessive computational time [17]. The results of the present paper indicate that the DTQ method achieves the same accuracy as a Fokker-Planck solver with less computational effort, further motivating the potential use of the DTQ method in many inference applications.

We now review alternative approaches and prior work related to either the general problem of computing the density of (1), or the particular case of the DTQ method.

## 1.1 Alternative Approaches

If the drift  $f$  and diffusion  $g$  are sufficiently smooth, then  $p$  satisfies the forward Kolmogorov (or Fokker-Planck) equation [27]:

$$\frac{\partial}{\partial t}p(x, t) = -\frac{\partial}{\partial x}[f(x, t)p(x, t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[g^2(x, t)p(x, t)]. \quad (2)$$

Prescribing an initial condition  $p(x, 0)$ , we may then solve (2) to obtain the density  $p(x, T)$  at time  $T > 0$ . The solution of (2) must satisfy the normalization condition  $\int_{x=-\infty}^{\infty} p(\cdot, t) dx = 1$ , which implies boundary conditions of the form  $\lim_{|x| \rightarrow \infty} p(x, t) = 0$ .

We view the DTQ method as an alternative to numerical methods for the solution of (2). The primary purpose of the present paper is to demonstrate intrinsic properties—both theoretical and empirical—of the DTQ method. We leave for future work an extensive comparison of the DTQ method against numerical methods for the solution of (2); such methods include finite difference, finite element, meshless, and Hermite spectral methods [24, 26, 10, 22]. Nevertheless, in the present work, we do compare the performance of the DTQ method against an elementary finite difference method applied to (2). The finite difference method we consider is first-order in time and second-order in space. For a particular test problem at the finest grid resolution we consider, the DTQ method computes a solution with  $L^1$  error  $\approx 3 \times 10^{-3}$  more than 100 times faster than our Fokker-Planck method.

Besides the numerical solution of (2), another method one might use to estimate the density of (1) involves sampling. Specifically, one can employ any convergent numerical method to step (1) forward in time from  $t = 0$  to  $t = T$ , thereby generating one sample of  $X_T$ . Repeating this procedure many times, one can obtain enough samples of  $X_T$  to compute a statistical estimate of the density at time  $T$ . For instance, one could compute a histogram or a kernel density estimate. Several methods in the literature can be viewed as special cases and extensions of this approach [16, 18, 23, 15]. In such methods, the accuracy of the density will be controlled by two parameters: the temporal step size and the number of sample paths. If there are  $N_S$  samples, then a typical stochastic time-stepping method will contribute an error of  $N_S^{-1/2}$  and kernel density estimation will contribute an error of, e.g.,  $N_S^{-4/5}$ . In comparison, the DTQ method's accuracy is also controlled by two parameters, the temporal step size and the spatial grid size. However, the spatial discretization using the trapezoidal rule on the real line will contribute an error that decays exponentially in the spatial grid size [32]. For this reason, we believe the DTQ method will be a strong alternative to a sampling-based method.

Returning to the forward Kolmogorov or Fokker-Planck equation (2), we see that smoothness of  $f$  and  $g$  is required in order to have classical solutions. The implementation of the DTQ method itself does not utilize derivatives (whether exact or approximate) of  $f$  and  $g$ . At the same time, the reader will note that our convergence theory assumes analyticity of  $f$  and  $g$  on a strip in the complex plane that contains the real line. We give two reasons for assuming analyticity. First, many models of scientific interest involve functions  $f$  and  $g$  that do satisfy these hypotheses. Second, in order to apply exponential error estimates for the trapezoidal rule [32], it is essential that our integrand, which depends on  $f$  and  $g$ , be analytic on a strip.

Ultimately, we do expect that the hypotheses in the present convergence proof can be relaxed. Let  $\tilde{p}$  be an approximate density that is computed in exactly the same way as  $\hat{p}$  except for truncation of the infinite domain/series. Our empirical results clearly show first-order convergence of  $\tilde{p}$  to  $p$ , even when not all of the hypotheses of our theorem are satisfied. Suppose that, inspired by these results, we discover how to prove convergence of  $\hat{p}$  to  $\tilde{p}$  assuming, for instance, that both  $f$  and  $g$  possess merely 4 bounded continuous derivatives. This will not immediately improve our ability to conclude that  $\hat{p}$  converges to  $p$ ; the existing result on convergence of  $\tilde{p}$  to  $p$  requires that both  $f$  and  $g$  are  $C^\infty$  with bounded derivatives of all orders [2]. To make true progress on the problem, we must relax the conditions of convergence for both  $\hat{p} \rightarrow \tilde{p}$  and  $\tilde{p} \rightarrow p$ . This is outside the scope of the present work.

## 1.2 Prior Work

The DTQ method that we propose in the present paper is an extension of our prior work on computing the density function for stochastic (delay) differential equations [5, 4, 6]. In fact, the method from [6], when adapted to equations with no time delay, is the method in the present paper. Our prior works did not address convergence from a theoretical standpoint, nor did they present empirical results of monotonic convergence that are in strict accordance with theory. The present paper addresses both of these issues.

When we derive the DTQ method, we make use of the fact that a time-discretization of (1) can be viewed as a discrete-time Markov chain on a continuous state space. Suppose we were to take a different point of view, that of trying to design a discrete-time Markov chain on a discrete state space whose law or density approximates well that of the original SDE. In this case, there are extensive results going back to the work of H. J. Kushner [20]. Like a discrete-time, discrete-time Markov chain, the DTQ algorithm can be written in the form  $\hat{p}(t_{n+1}) = A\hat{p}(t_n)$ , where  $A$  is a matrix (possibly with an infinite number of rows and columns) and  $\hat{p}(t_j)$  represents the approximate density at time  $t_j$ . However, because of the quadrature-based derivation of the DTQ algorithm, the matrix  $A$  is, in general, not a Markov transition matrix. We find it both mathematically interesting and practically useful that, in spite of this, the DTQ method's  $\hat{p}$  converges exponentially to  $\tilde{p}$ .

The Chapman-Kolmogorov equation that is at the center of this paper—see (6)—has appeared in other papers [25, 29]. In these works, the right-hand side of the Chapman-Kolmogorov equation is interpreted as an expected value that can be computed stochastically, i.e., using Monte Carlo methods. In our approach, we use deterministic quadrature to evaluate the right-hand side of the Chapman-Kolmogorov equation. There is only one prior paper we found that features this approach, albeit in a different context, that of a nonlinear autoregressive time series model [9]. The convergence results in [9] are of a different nature than ours, because they involve taking the continuum limit in space but *not* in time. In the present work, we are interested in the error made by the DTQ method as both the temporal and spatial grid spacings vanish.

## 1.3 Summary of Results and Outline

The main result of this paper is a provably convergent method for computing an approximation  $\hat{p}$  of the density  $p$  for the SDE (1). Let  $h > 0$  and  $k > 0$  denote, respectively, the temporal and spatial step sizes. Assume that  $k \propto h^\rho$  for  $\rho > 1/2$ , and assume that  $f$  and  $g$  are sufficiently regular (more precisely, admissible in the sense of Definition 2). Under these conditions, in Sections 4 and 5, we prove that  $\hat{p}$  converges to  $\tilde{p}$  in  $L^1$ , and that the error decays exponentially in  $h$ .

Specifically, there exists a constant  $r > 0$  such that the leading order  $L^1$  error term is proportional to  $h^{-1} \exp(-rh^{1/2-\rho})$ —see Theorem 2. As a consequence of this result and the results of [2], we conclude that  $\hat{p}$  converges to  $p$  in  $L^1$ , and that the error decays linearly with  $h$ —see Corollary 1.

Up to and including Section 5, our results pertain to an idealized version of the DTQ algorithm in which we track the density  $\hat{p}$  at an infinite number of discrete grid points. In Section 6, we study the effect of boundary truncation. Our main tool in this section is a Chernoff bound on the tail sum of  $\hat{p}$  that we establish through the moment generating function. Let  $\mathring{p}$  denote the approximation of  $\hat{p}$  obtained by summing over precisely  $2M + 1$  grid points from  $-y_M = -Mk$  to  $y_M = Mk$ . The quantity  $\mathring{p}$  is what we actually compute when we implement the DTQ method. In Lemma 9, we show that if  $y_M \rightarrow \infty$  at a logarithmic rate, i.e.,  $y_M \propto \log h^{-1}$ , then the  $L^1$  error between  $\mathring{p}$  and  $\hat{p}$  is  $O(h)$ . Combining this with our earlier results, this establishes  $L^1$  convergence of  $\mathring{p}$  to the true density  $p$ —see Corollary 2.

In Section 7, we study the performance of the DTQ method. For a suite of six test problems for which we have access to the exact solution, our numerical tests confirm  $O(h)$  convergence of  $\mathring{p}$  to  $p$ . This remains true for drift  $f$  and diffusion functions  $g$  that do not strictly satisfy the hypotheses of our convergence theory. We also present a finite difference method for solving (2); we compare this method against three slightly different implementations of the DTQ method. The comparison indicates that the DTQ method—which we believe is being analyzed here for the first time—is competitive with standard numerical methods for (2).

Before proceeding, we give a more detailed derivation of the DTQ method in Section 2 and then introduce necessary assumptions and notation in Section 3.

## 2 Problem Setup

We begin with a more detailed derivation of the DTQ method. First, we discretize (1) in time using the explicit Euler-Maruyama method:

$$x_{n+1} = x_n + f(x_n)h + g(x_n)\sqrt{h}Z_{n+1}, \quad (3)$$

where  $h > 0$  is a fixed time step and  $Z_{n+1}$  is a random variable with a standard (mean zero, variance one) Gaussian distribution. We let  $\tilde{p}(x, t_n)$  denote the probability density function of  $x_n$ . Note that this differs from  $p(x, t_n)$ .

From (3), we observe that the density of  $x_{n+1}$  given  $x_n = y$  is Gaussian with mean  $y + f(y)h$  and variance  $hg^2(y)$ . Let us denote this conditional density by  $p_{n+1,n}(x|y)$ ; then

$$p_{n+1,n}(x|y) = G(x, y) := \frac{1}{\sqrt{2\pi g^2(y)h}} \exp\left(-\frac{(x - y - f(y)h)^2}{2g^2(y)h}\right). \quad (4)$$

Note that, for any  $y \in \mathbb{R}$ ,

$$\int_{x=-\infty}^{\infty} G(x, y) dx = 1. \quad (5)$$

Applying this to (3), we obtain the following evolution equation:

$$\tilde{p}(x, t_{n+1}) = \int_{-\infty}^{\infty} p_{n+1,n}(x|y) \tilde{p}(y, t_n) dy. \quad (6)$$

This is the Chapman-Kolmogorov equation for the discrete-time, continuous-space Markov chain given by (3). Similar equations are often employed in the literature on inference for diffusions—see [25], [29], [14, Chap 6.3.3], and [19].

Let us define an equispaced temporal grid by  $t_n = nh$  with  $h = T/N$ . In principle, we can now repeatedly apply (6) to determine  $\tilde{p}(x, T)$ . This assumes we can perform the integral over the real line.

To compute (6) in practice, we use numerical quadrature. Here we employ the trapezoidal rule, enabling us to make use of exponential error estimates [32, 31, 21]. To begin with, we apply the trapezoidal rule on the real line. Later, we explain how to incorporate the effects of a finite, truncated integration domain.

Assume the domain  $\mathbb{R}$  is discretized via an equispaced grid  $y_j = jk$  where  $k > 0$  is fixed. Then our discrete-time, discrete-space evolution equation is

$$\hat{p}(x, t_{n+1}) = k \sum_{j=-\infty}^{\infty} G(x, y_j) \hat{p}(y_j, t_n). \quad (7)$$

Except for the fact that we have not yet truncated the infinite sum, this is the DTQ method.

Thus far we have avoided the discussion of initial conditions for both  $\tilde{p}$  and  $\hat{p}$ . For the purposes of exposition, we assume a constant initial condition  $X_0 = C$ , which implies  $p(x, 0) = \tilde{p}(x, 0) = \delta(x - C)$ . This choice is not essential to either the use or convergence of the DTQ method. In fact, the choice of a point mass initial condition requires special handling, because we cannot discretize  $\tilde{p}(x, 0)$  directly. We insert  $n = 0$  into (6), use  $\tilde{p}(x, 0) = \delta(x - C)$ , and obtain the non-singular initial condition

$$\hat{p}(x, t_1) = \tilde{p}(x, t_1) = G(x, C). \quad (8)$$

This enables us to initialize and iteratively use both (6) and (7) for  $n \geq 1$ .

Our main task in Sections 4 and 5 is to estimate  $\|\hat{p}(\cdot, T) - \tilde{p}(\cdot, T)\|_1$ . Before we start the proof of Theorem 2, we introduce necessary notation and assumptions.

### 3 Notation and Assumptions

We will use the Roman  $i$  for the imaginary unit ( $i = \sqrt{-1}$ ) and reserve the Italic  $i$  for an index of summation. We denote the  $L^1$  norm of a function  $f : \mathbb{R} \rightarrow \mathbb{R}$  by

$$\|f\|_1 = \int_{-\infty}^{\infty} |f(x)| dx.$$

We denote the  $\ell^1$  norm of the sequence  $\{z_j\}_{j=-\infty}^{\infty}$  by

$$\|z\|_{\ell^1} = \sum_{j=-\infty}^{\infty} |z_j|.$$

For a function  $f : \mathbb{R} \rightarrow \mathbb{R}$ , we understand  $\|f\|_{\ell^1}$  to be the norm of the sequence obtained by applying  $f$  on the spatial grid:

$$\|f\|_{\ell^1} = \sum_{j=-\infty}^{\infty} |f(jk)|,$$

where again  $k > 0$  denotes the grid spacing. We use  $\lceil x \rceil$  to denote the smallest integer greater than or equal to  $x$ , and  $\lfloor x \rfloor$  to denote the largest integer less than or equal to  $x$ . The following definition is from the literature [21].

**Definition 1.** For  $a > 0$ , let  $S_a$  denote the infinite strip of width  $2a$  given by

$$S_a = \{z \in \mathbb{C} : |\Im(z)| < a\}.$$

Then  $B(S_a)$  is the set of functions such that  $\varphi \in B(S_a)$  iff  $\varphi$  is analytic in  $S_a$ ,

$$\int_{-a}^a |\varphi(x + iy)| dy = O(|x|^\alpha), \quad x \rightarrow \pm\infty, \quad 0 \leq \alpha < 1, \quad (9)$$

and

$$\mathcal{N}(\varphi, S_a) \equiv \lim_{y \rightarrow a^-} \left\{ \int_{-\infty}^{\infty} |\varphi(x + iy)| dx + \int_{-\infty}^{\infty} |\varphi(x - iy)| dx \right\} < \infty. \quad (10)$$

The next definition encapsulates the constraints that the coefficient functions  $f$  and  $g$  in the original SDE (1) must satisfy in order for us to show exponential convergence of  $\hat{p}$  to  $\tilde{p}$ .

**Definition 2.** In this paper, we say that  $f$  and  $g$  are admissible if they satisfy the following properties. First, there exists  $d > 0$  such that  $f$  and  $g$  are analytic on the strip  $S_d$ . Additionally, there exist positive, finite, real constants  $M_1$ ,  $M_2$ ,  $M_3$ , and  $M_4$  such that for all  $z \in S_d$ ,

$$|f'(z)| \leq M_1 \quad (11a)$$

$$M_2 \leq |g(z)| \leq M_3 \quad (11b)$$

$$\Re(g(z)) \neq 0 \quad (11c)$$

$$|g'(z)| \leq M_4. \quad (11d)$$

We now state a theorem that gives an exponential error estimate for the trapezoidal rule [21], one that we shall use to bound the error made in one step of the DTQ method. Other error estimates can be found in the literature [31, 32].

**Theorem 1.** Suppose  $\varphi \in B(S_d)$  and  $k > 0$ . Let

$$\eta = \int_{-\infty}^{\infty} \varphi(x) dx - k \sum_{j=-\infty}^{\infty} \varphi(jk).$$

Then

$$|\eta| \leq \frac{\mathcal{N}(\varphi, S_d)}{2 \sinh(\pi d/k)} \exp(-\pi d/k).$$

*Proof.* See [21, Theorem 2.20]. □

## 4 Preliminary Theory

In this section, we prove several lemmas that are essential ingredients for the convergence theorem in Section 5. The overall goal of these lemmas is to show that the integrand

$$\varphi(x, y, t_n) = G(x, y) \hat{p}(y, t_n), \quad (12)$$

considered as a function of  $y$  for the purposes of quadrature, satisfies the hypotheses of Theorem 1.

The first lemma enables us to pass from an estimate of the error made in one time step to an estimate of the error made across a non-zero interval of time, even as the number of time steps becomes infinite.

**Lemma 1.** *Suppose that  $\xi(h) \geq 0$  satisfies  $\lim_{h \rightarrow 0^+} \xi(h) = 0$ . Suppose there exist  $\gamma > 1$ ,  $\epsilon > 0$  and  $h_0 > 0$  such that  $\xi(h) \leq \epsilon h^\gamma$  for all  $h < h_0$ . Fix  $T > 0$ ,  $N \in \mathbb{N}^+$ , and let  $h = T/N$ . Then*

$$\lim_{N \rightarrow \infty} \left[ h \sum_{j=0}^{N-1} (1 + \xi(h))^j \right] = T.$$

*Proof.* Take  $N$  sufficiently large so that  $h < 1$  and  $h < h_0$ . Then we calculate

$$\begin{aligned} \sum_{j=0}^{N-1} (1 + \xi(h))^j &= \xi(h)^{-1} [(1 + \xi(h))^N - 1] = \sum_{j=1}^N \binom{N}{j} \xi(h)^{j-1} \\ &\leq \frac{T}{h} + \sum_{j=2}^N \frac{T^j \epsilon^{j-1}}{j!} h^{\gamma(j-1)-j} \end{aligned}$$

Using  $h < 1$ , we have

$$h \sum_{j=0}^{N-1} (1 + \xi(h))^j \leq T + \sum_{j=2}^N \frac{T^j \epsilon^{j-1}}{j!} h^{(\gamma-1)(j-1)} \leq T + \epsilon^{-1} h^{\gamma-1} \exp(T\epsilon).$$

We have shown that the limit is  $T$ , and that the correction term to the limit is  $O(h^{\gamma-1})$ .  $\square$

Next, we estimate the  $\ell^1$  norm of the discrete Gaussian. This estimate is standard, but we include it here for the sake of completeness.

**Lemma 2.** *For all  $y \in \mathbb{R}$  and all  $h, k > 0$ ,*

$$k \|G(\cdot, y)\|_{\ell^1} \leq 1 + 4 \exp \left( -\frac{2\pi^2 g^2(y)h}{k^2} \right). \quad (13)$$

*Proof.* Let

$$\phi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x - \mu)^2}{2\sigma^2} \right). \quad (14)$$

Note that for any  $d > 0$ , on the strip  $S_d$ ,  $\phi$  satisfies the hypotheses of Theorem 1. In particular,

$$\int_{-\infty}^{\infty} \left| \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left( -\frac{(x + id - \mu)^2}{2\sigma^2} \right) \right| dx = \exp \left( \frac{d^2}{2\sigma^2} \right).$$

As the right-hand side does not change when we replace  $d$  by  $-d$ , we have  $\mathcal{N}(\phi, S_d) = 2 \exp(d^2/(2\sigma^2))$ . Therefore, applying Theorem 1,

$$\left| \int_{-\infty}^{\infty} \phi(x) dx - k \sum_{j=-\infty}^{\infty} \phi(jk) \right| \leq \frac{\exp(d^2/(2\sigma^2))}{\sinh(\pi d/k)} \exp \left( -\frac{\pi d}{k} \right) \leq 4 \exp \left( \frac{d^2}{2\sigma^2} - \frac{2\pi d}{k} \right),$$



where we have used  $(\sinh(\pi d/k))^{-1} \leq 4 \exp(-\pi d/k)$ . The right-hand side is minimized at  $d = 2\pi\sigma^2/k$ . Also,  $\int_{-\infty}^{\infty} \phi(x) dx = 1$ . Hence

$$k \sum_{j=-\infty}^{\infty} \phi(jk) \leq 1 + 4 \exp\left(-\frac{2\pi^2\sigma^2}{k^2}\right). \quad (15)$$

Note that  $\phi(x) = G(x, y)$  with  $\mu = y + f(y)h$  and  $\sigma^2 = g^2(y)h$ . Then (15) is (13).  $\square$

For each  $t_n$ , we think of  $\{\hat{p}(x_j, t_n)\}_{j=-\infty}^{\infty}$  as an infinite sequence. It is important to estimate the  $\ell^1$  norm of this sequence.

**Lemma 3.** *If  $g$  is admissible in the sense of Definition 2, then for all  $h, k > 0$ ,*

$$\|\hat{p}(\cdot, t_{n+1})\|_{\ell^1} \leq \|\hat{p}(\cdot, t_1)\|_{\ell^1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^n. \quad (16)$$

*Proof.* We begin by evaluating (7) at  $x = x_i$ :

$$\hat{p}(x_i, t_{n+1}) = k \sum_{j=-\infty}^{\infty} G(x_i, y_j) \hat{p}(y_j, t_n). \quad (17)$$

Before proceeding, let us discuss the convergence of the infinite series on the right-hand side for fixed  $h$  and  $k$ . Using (11b), we have for  $G$  the elementary bound  $0 \leq G(x, y) \leq (2\pi M_2^2 h)^{-1/2}$ . Note that (8) and (13) together give us an  $\ell^1$  bound on  $\{\hat{p}(jk, t_1)\}_{j=-\infty}^{\infty}$ . Combining these two bounds, it is clear that (17) converges for  $n = 1$ .

Now, as an induction hypothesis, assume that for a particular  $n \geq 1$ , we have  $\hat{p}(y, t_n) \geq 0$  and that  $\|\hat{p}(\cdot, t_n)\|_{\ell^1} < \infty$ . We will establish an  $\ell^1$  bound for  $\hat{p}(\cdot, t_{n+1})$ .

By the induction hypothesis, we know that the infinite series on the right-hand side of (17) converges. We see that all terms in the infinite series are nonnegative, so  $\hat{p}(y, t_{n+1}) \geq 0$ . Additionally, both sides of (17) do not change upon taking absolute values. We sum over all  $i$  and interchange the order of summation—this is justified because, again, all terms are nonnegative. We obtain

$$\|\hat{p}(\cdot, t_{n+1})\|_{\ell^1} = \sum_{j=-\infty}^{\infty} \left[ k \sum_{i=-\infty}^{\infty} G(x_i, y_j) \right] \hat{p}(y_j, t_n).$$

Applying (13) and (11b), we have

$$\|\hat{p}(\cdot, t_{n+1})\|_{\ell^1} \leq (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2)) \|\hat{p}(\cdot, t_n)\|_{\ell^1}. \quad (18)$$

This shows that  $\|\hat{p}(\cdot, t_{n+1})\|_{\ell^1} < \infty$ , finishing the induction step. Combining this with the elementary bound on  $G$ , it is clear that the series on the right-hand side of (17) converges for all  $n \geq 1$ . This implies the convergence of (7), as an infinite series, for all  $n \geq 1$ .

Iterating the inequality (18)  $n$  times, we derive (16).  $\square$

The importance of Lemma 3 is that it enables us to give asymptotic conditions on  $h$  and  $k$  such that  $\hat{p}$  is normalized correctly.

**Lemma 4.** *Suppose that  $g$  is admissible in the sense of Definition 2, and that  $k = r_1 h^\rho$  for constants  $r_1 > 0$  and  $\rho > 1/2$ . Assume that  $N = T/h$  for some fixed  $T > 0$ . Then for  $1 \leq n \leq N + 1$ ,*

$$\lim_{h \rightarrow 0} k \|\hat{p}(\cdot, t_n)\|_{\ell^1} = 1. \quad (19)$$

*Proof.* Applying the hypotheses to the exponential term in (16) with  $n = N = T/h$ , we have

$$\lim_{h \rightarrow 0} (1 + 4 \exp(-2\pi^2 M_2^2 r_1^{-2} h^{-2\rho+1}))^{T/h} = 1. \quad (20)$$

For any  $n \in \{0, 1, \dots, N\}$ , we have

$$\lim_{h \rightarrow 0} \|\hat{p}(\cdot, t_{n+1})\|_{\ell^1} / \|\hat{p}(\cdot, t_1)\|_{\ell^1} = 1. \quad (21)$$

Next, we combine the fact that  $\hat{p}(\cdot, t_1)$  is Gaussian with (13) to conclude that  $k\|\hat{p}(\cdot, t_1)\|_{\ell^1} \rightarrow 1$  as  $k \rightarrow 0$ . Then (19) follows immediately from (21).  $\square$

**Lemma 5.** *Suppose that  $f$  and  $g$  are admissible in the sense of Definition 2, and that  $a < \min\{d, M_2^2/(2M_3M_4)\}$ . Then for any  $x, y \in \mathbb{R}$ , there exists  $A_2 > 0$  such that*

$$|G(x, y + ia)| = \frac{1}{\sqrt{2\pi h |g(y + ia)|^2}} \exp\left(-\frac{A_2 x^2 + A_1 x + A_0}{4|g(y + ia)|^4 h}\right), \quad (22)$$

and there exists  $\gamma_0 \in (0, 2)$  such that

$$|G(x, y + ia)| \leq \frac{1}{\sqrt{2\pi h M_2^2}} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right).$$

*Proof.* We obtain (22) by direct calculation of  $|G(x, y + ia)|$ . The coefficients  $A_2$ ,  $A_1$ , and  $A_0$  are defined by

$$A_2 = g^2(y - ia) + \text{c.c.} \quad (23a)$$

$$A_1 = -2g^2(y - ia)(y + ia + f(y + ia)h) + \text{c.c.} \quad (23b)$$

$$A_0 = g^2(y - ia)(y^2 - a^2 + f^2(y + ia)h^2 + 2yia + 2(y + ia)f(y + ia)h) + \text{c.c.} \quad (23c)$$

By “c.c.” we mean the complex conjugate of the preceding term. We have used the fact that because  $f$  and  $g$  are analytic on  $S_d$ , and because they are real-valued when restricted to the real axis, both  $f$  and  $g$  commute with complex conjugation. That is,  $\overline{f(y + ia)} = f(y - ia)$  and similarly for  $g$  and  $g^2$ . The upshot is that  $A_2$ ,  $A_1$ , and  $A_0$  are all real.

Let us now prove that  $A_2 > 0$ . Define the function

$$\theta(y, \epsilon) = g^2(y - i\epsilon) + g^2(y + i\epsilon),$$

for  $\epsilon \in [0, d]$ . For each fixed  $y$ , by the mean-value theorem, there exists  $\xi$  such that

$$\theta(y, \epsilon) - \theta(y, 0) = \epsilon \frac{\partial \theta}{\partial \epsilon}(y, \xi).$$

Note that  $\xi$  may depend on  $\epsilon$  and  $y$ . Now we use (11) to compute

$$\sup_{y \in \mathbb{R}, \epsilon \in (-d, d)} \left| \frac{\partial \theta}{\partial \epsilon} \right| = 4 \sup_{\substack{y \in \mathbb{R} \\ \epsilon \in (-d, d)}} |\Im(g(y + i\epsilon)g'(y + i\epsilon))| \leq 4M_3M_4. \quad (24)$$

Then using the previous two equations together with (11b), we have

$$\theta(y, \epsilon) \geq \theta(y, 0) - 4\epsilon M_3M_4 \geq 2M_2^2 - 4\epsilon M_3M_4. \quad (25)$$

The right-hand side will be positive as long as  $\epsilon < \min\{d, M_2^2/(2M_3M_4)\}$ . Given the hypothesis on  $a$  in the statement of the lemma,  $\theta(y, a) = A_2$  will be positive.

Because  $A_2 > 0$ , we can maximize the right-hand side of (22) as a function of  $x$ —the global maximum occurs at  $x = -A_1/(2A_2)$ . Then we have

$$|G(x, y + ia)| \leq \frac{1}{\sqrt{2\pi h M_2^2}} \exp \left( \frac{(2a + ih(f(y - ia) - f(y + ia)))^2}{4h(g^2(y + ia) + g^2(y - ia))} \right)$$

We suppose that  $a = bM_2^2/(2M_3M_4)$  for some  $b \in (0, 1)$  such that  $a < d$ . Then the lower bound (25) implies  $\theta(y, a) \geq 2M_2^2(1 - b)$ . We define  $\gamma_0 = 2(1 - b) \in (0, 2)$  and write

$$|G(x, y + ia)| \leq \frac{1}{\sqrt{2\pi h M_2^2}} \exp \left( \frac{(2a + ih(f(y - ia) - f(y + ia)))^2}{h\gamma_0 M_2^2} \right) \quad (26)$$

Let  $\Gamma$  be the segment connecting  $y - ia$  to  $y + ia$ . Note that  $a < d$  implies that  $\Gamma$  is completely contained in the strip  $S_d$  where  $f$  is analytic. Using (11a), we have

$$\begin{aligned} |2a + ih(f(y - ia) - f(y + ia))| &\leq 2|a| + h|f(y + ia) - f(y - ia)| \\ &\leq 2|a| + h \left| \oint_{\Gamma} f'(z) dz \right| \\ &\leq 2|a| + h \oint_{\Gamma} |f'(z)| |dz| \\ &\leq 2|a|(1 + hM_1) \end{aligned}$$

Using this estimate in (26) finishes the proof.  $\square$

**Lemma 6.** *Suppose that  $f$  and  $g$  are admissible in the sense of Definition 2, and that  $a < \min\{d, M_2^2/(2M_3M_4)\}$ . Then the integrand (12), considered as a function of  $y$ , is a member of  $B(S_a)$ , i.e.,  $\varphi(x, \cdot, t_n) \in B(S_a)$ .*

*Proof.* There are three conditions for membership in  $B(S_a)$ , which we verify in turn. First, it is simple to check that  $\varphi$  is analytic on  $S_a$ ; this follows naturally from (11c) and the lower bound in (11b).

At time step  $t_1$ , we have  $\hat{p}(y, t_1) = G(y, C)$ , which is analytic. The arguments made earlier regarding the convergence of (17) hold equally well with  $x_i$  replaced by any  $x$ . This implies that for  $n \geq 1$ ,  $\hat{p}(y, t_{n+1})$  is analytic in  $y$  on  $S_d$ , so the integrand  $\varphi$  is analytic on  $S_a \subset S_d$ .

Next, we consider

$$\Phi(x, y, t_n) = \int_{b=-a}^a |\varphi(x, y + ib, t_n)| db. \quad (27)$$

Since

$$\hat{p}(y + ia, t_{n+1}) = k \sum_{j=-\infty}^{\infty} G(y + ia, z_j) \hat{p}(z_j, t_n), \quad (28)$$

we have

$$\begin{aligned}
\Phi(x, y, t_{n+1}) &\leq k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{b=-a}^a |G(y + ib, z_j)| |G(x, y + ib)| db \\
&= k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) G(y, z_j) \int_{b=-a}^a \exp\left(\frac{b^2}{2g^2(z_j)h}\right) |G(x, y + ib)| db \\
&\leq \frac{1}{\sqrt{2\pi h M_2^2}} \int_{b=-a}^a \exp\left(\frac{b^2}{2M_2^2 h}\right) \exp\left(\frac{b^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right) db \\
&\quad \times k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) G(y, z_j).
\end{aligned}$$

To arrive at the last line, we have applied Lemma 5 and (11b). There is only one remaining term on the right-hand side that depends on  $y$ . As  $|y| \rightarrow \infty$ , we have that  $G(y, z_j) \rightarrow 0$ . So, as  $|y| \rightarrow \infty$ , we have that  $\Phi(x, y, t_{n+1}) = O(|y|^\alpha)$  for  $\alpha = 0$ , satisfying (9).

Next, we establish a bounded, real function  $L_n$  such that for each  $x \in \mathbb{R}$ ,

$$\begin{aligned}
\mathcal{N} &:= \int_{y=-\infty}^{\infty} |G(x, y + ia) \hat{p}(y + ia, t_n)| dy \\
&\quad + \int_{y=-\infty}^{\infty} |G(x, y - ia) \hat{p}(y - ia, t_n)| dy \leq L_n(x) < \infty. \quad (29)
\end{aligned}$$

We need this estimate in order to apply Theorem 1. For this purpose, we seek an upper bound on  $\mathcal{N}$  that does not depend essentially on the spatial discretization parameter  $k$ . Starting again from (28), we have

$$\begin{aligned}
&\int_{y=-\infty}^{\infty} |G(x, y + ia) \hat{p}(y + ia, t_{n+1})| dy \\
&\leq k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{y=-\infty}^{\infty} |G(y + ia, z_j)| |G(x, y + ia)| dy \\
&= k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{y=-\infty}^{\infty} \exp\left(\frac{a^2}{2g^2(z_j)h}\right) G(y, z_j) |G(x, y + ia)| dy \\
&\leq \exp\left(\frac{a^2}{2M_2^2 h}\right) \left[ k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{y=-\infty}^{\infty} G(y, z_j) |G(x, y + ia)| dy \right] \quad (30)
\end{aligned}$$

$$\begin{aligned}
&\leq k \exp\left(\frac{a^2}{2M_2^2 h}\right) \|\hat{p}(\cdot, t_n)\|_{\ell^1} \sup_z \left[ \int_{y=-\infty}^{\infty} G(y, z) |G(x, y + ia)| dy \right] \\
&\leq k \exp\left(\frac{a^2}{2M_2^2 h}\right) \|\hat{p}(\cdot, t_n)\|_{\ell^1} \psi(x, a), \quad (31)
\end{aligned}$$

where

$$\psi(x, a) = \sup_{z \in \mathbb{R}} \left[ \int_{y=-\infty}^{\infty} G(y, z) |G(x, y + ia)| dy \right]. \quad (32)$$

Examining (22), we see that the right-hand side of (31) is invariant under the reflection  $a \mapsto (-a)$ . We define the real-valued function

$$L_{n+1}(x) = 2k \exp\left(\frac{a^2}{2M_2^2 h}\right) \|\hat{p}(\cdot, t_n)\|_{\ell^1} \psi(x, a),$$

and note that (31) implies  $\mathcal{N} \leq L_n(x)$ , as required by (29). Our task now is to demonstrate that  $L_n$  is finite. By Lemma 5 and (5), we have

$$\begin{aligned} \psi(x, a) &\leq \sup_z \left[ \frac{1}{\sqrt{2\pi h M_2^2}} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right) \int_{y=-\infty}^{\infty} G(y, z) dy \right] \\ &\leq \frac{1}{\sqrt{2\pi h M_2^2}} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right). \end{aligned}$$

Using this estimate in (31), we obtain

$$L_{n+1}(x) \leq 2k \exp\left(\frac{a^2}{2M_2^2 h}\right) \|\hat{p}(\cdot, t_n)\|_{\ell^1} \frac{1}{\sqrt{2\pi h M_2^2}} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right).$$

Note that the bound on the right-hand side does not depend on  $x$  at all. The dependence on  $k$  is confined to the terms  $k\|\hat{p}(\cdot, t_n)\|$ . By Lemmas 2 and 3 together with (13),

$$k\|\hat{p}(\cdot, t_n)\| \leq (1 + 4\exp(-2\pi^2 g^2(C)h/k^2)) (1 + 4\exp(-2\pi^2 M_2^2 h/k^2))^{n-1} \leq 5^n < \infty$$

for all  $k \geq 0$ . In sum, we have shown that for fixed  $h > 0$ , fixed  $n \geq 1$ , and  $a < \min\{d, M_2^2/(2M_3M_4)\}$ ,  $L_n(x)$  is bounded uniformly in  $x$  and  $k$ . We have demonstrated that (29) holds. We conclude that  $\varphi(x, \cdot, t_n) \in B(S_a)$ .  $\square$

## 5 Convergence Theorem

Let

$$E(y, t_n) = \tilde{p}(y, t_n) - \hat{p}(y, t_n). \quad (33)$$

In this section, we establish conditions under which  $\|E(\cdot, T)\|_1$  goes to zero at an exponential rate.

**Theorem 2.** *Assume that  $f$  and  $g$  are admissible in the sense of Definition 2. Assume that*

$$k = r_1 h^\rho \quad (34)$$

*for constants  $r_1 > 0$  and  $\rho > 1/2$ . Choose  $a < \min\{d, M_2^2/(2M_3M_4)\}$  such that*

$$a = r_2 h^{1/2} \quad (35)$$

*for some  $r_2 > 0$ . For fixed  $T > 0$ , choose*

$$h \in (0, \min\{T, (M_2^2/(4M_3M_4r_2))^2\}) \quad (36)$$

*such that  $N = T/h \in \mathbb{N}^+$ . To be clear,  $r_1$  and  $r_2$  are constants that do not depend on  $h$ . Then*

$$\|E(\cdot, T)\|_1 \leq c_\star h^{-1} \exp(-2\pi r_2 r_1^{-1} h^{1/2-\rho})(1 + o(h) + o(k)) \quad (37)$$

*where  $o(h)$  and  $o(k)$  stand for terms that vanish as  $h \rightarrow 0$  and  $k \rightarrow 0$ , and  $c_\star > 0$  is a constant that does not depend on  $h$ .*

*Proof.* We begin with

$$\begin{aligned}\tilde{p}(x, t_{n+1}) &= \int_{y=-\infty}^{\infty} G(x, y) \tilde{p}(y, t_n) dy \\ &= \int_{y=-\infty}^{\infty} G(x, y) \hat{p}(y, t_n) dy + \int_{y=-\infty}^{\infty} G(x, y) E(y, t_n) dy.\end{aligned}$$

We now apply the trapezoidal rule to the first integral. For each  $x$  and  $t_n$ , we let  $\tau(x, t_n)$  denote the quadrature error incurred, i.e.,

$$\begin{aligned}\int_{y=-\infty}^{\infty} G(x, y) \hat{p}(y, t_n) dy &= k \sum_{j=-\infty}^{\infty} G(x, y_j) \hat{p}(y_j, t_n) + \tau(x, t_n) \\ &= \hat{p}(x, t_{n+1}) + \tau(x, t_n).\end{aligned}\tag{38}$$

We use this in the previous equation to derive

$$E(x, t_{n+1}) = \int_{y=-\infty}^{\infty} G(x, y) E(y, t_n) dy + \tau(x, t_n).$$

Taking absolute values, we apply the triangle inequality together with  $G \geq 0$  to obtain

$$|E(x, t_{n+1})| \leq \int_{y=-\infty}^{\infty} G(x, y) |E(y, t_n)| dy + |\tau(x, t_n)|.$$

Integrating over  $x$  and using Fubini's theorem and (5), we have

$$\|E(\cdot, t_{n+1})\|_1 - \|E(\cdot, t_n)\|_1 \leq \|\tau(\cdot, t_n)\|_1.\tag{39}$$

Summing both sides from  $n = 1$  to  $n = N - 1$  and using (8), we have

$$\|E(\cdot, T)\|_1 \leq \sum_{n=1}^{N-1} \|\tau(\cdot, t_n)\|_1.\tag{40}$$

We apply Lemma 6 and Theorem 1 to produce the estimate

$$|\tau(x, t_n)| \leq \frac{\mathcal{N}}{2 \sinh(\pi a/k)} \exp(-\pi a/k)\tag{41}$$

where  $\tau$  and  $\mathcal{N}$  are defined by (38) and (29), respectively. Combining (30) with (22), we have

$$\begin{aligned}\int_{y=-\infty}^{\infty} |G(x, y + ia) \hat{p}(y + ia, t_{n+1})| dy &\leq \exp\left(\frac{a^2}{2M_2^2 h}\right) \\ &\times k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{y=-\infty}^{\infty} \frac{G(y, z_j)}{\sqrt{2\pi h |g(y + ia)|^2}} \exp\left(-\frac{A_2 x^2 + A_1 x + A_0}{4|g(y + ia)|^4 h}\right) dy,\end{aligned}$$

where again  $A_2$ ,  $A_1$ , and  $A_0$  are defined by (23). We see that the right-hand side of this inequality is invariant under  $a \mapsto -a$ , and so we write

$$\mathcal{N} \leq 2 \exp\left(\frac{a^2}{2M_2^2 h}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \int_{y=-\infty}^{\infty} \frac{G(y, z_j)}{\sqrt{2\pi h |g(y + ia)|^2}} \exp\left(-\frac{A_2 x^2 + A_1 x + A_0}{4|g(y + ia)|^4 h}\right) dy.$$

For  $a < \min\{d, M_2^2/(2M_3M_4)\}$ , we have shown that the coefficient  $A_2$  is positive on  $S_a$ . This enables us to integrate both sides with respect to  $x$ :

$$\begin{aligned} \int_{x=-\infty}^{\infty} \mathcal{N} dx &\leq 2\sqrt{2} \exp\left(\frac{a^2}{2M_2^2h}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n) \\ &\quad \times \int_{y=-\infty}^{\infty} \frac{G(y, z_j) |g(y + ia)|}{\sqrt{g^2(y + ia) + g^2(y - ia)}} \exp\left(\frac{(2a + ih(f(y - ia) - f(y + ia)))^2}{4h(g^2(y + ia) + g^2(y - ia))}\right) dy. \end{aligned}$$

On the right-hand side, we have carried out the  $x$  integral first; the changing of the order of summation and integration is justified by the nonnegativity of every term. Next, we apply estimates established in the proof of Lemma 5. We obtain

$$\int_{x=-\infty}^{\infty} \mathcal{N} dx \leq 2\sqrt{2} \exp\left(\frac{a^2}{2M_2^2h}\right) \frac{M_3}{\gamma_0^{1/2} M_2} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n)$$

Combining this with (41), we have

$$\begin{aligned} \int_{x=-\infty}^{\infty} |\tau(x, t_n)| dx &\leq \\ &4\sqrt{2} \exp\left(\frac{a^2}{2M_2^2h}\right) \frac{M_3}{\gamma_0^{1/2} M_2} \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right) \exp(-2\pi a/k) k \sum_{j=-\infty}^{\infty} \hat{p}(z_j, t_n). \end{aligned}$$

Using (16), we obtain

$$\begin{aligned} \|\tau(\cdot, t_n)\|_1 &\leq 4\sqrt{2} M_3 \gamma_0^{-1/2} M_2^{-1} \exp\left(\frac{a^2}{2M_2^2h}\right) \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right) \exp(-2\pi a/k) \\ &\quad \times k \|\hat{p}(\cdot, t_1)\|_{\ell^1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-1}. \end{aligned}$$

We sum both sides from  $n = 1$  to  $n = N - 1$ :

$$\begin{aligned} \sum_{n=1}^{N-1} \|\tau(\cdot, t_n)\|_1 &\leq \sqrt{2} M_3 \gamma_0^{-1/2} M_2^{-1} \exp\left(\frac{a^2}{2M_2^2h}\right) \exp\left(\frac{a^2(1 + hM_1)^2}{h\gamma_0 M_2^2}\right) \\ &\quad \times h^{-1} \exp(-2\pi a/k) k \|\hat{p}(\cdot, t_1)\|_{\ell^1} \left[ h \sum_{n=1}^{N-1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-1} \right]. \quad (42) \end{aligned}$$

We now use (40) and hypotheses (34) and (35):

$$\begin{aligned} \|E(\cdot, T)\|_1 &\leq \sqrt{2} M_3 \gamma_0^{-1/2} M_2^{-1} \exp\left(\frac{r_2^2}{2M_2^2}\right) \exp\left(\frac{r_2^2(1 + hM_1)^2}{\gamma_0 M_2^2}\right) T \\ &\quad \times h^{-1} \exp(-2\pi r_2 r_1^{-1} h^{1/2-\rho}) k \|\hat{p}(\cdot, t_1)\|_{\ell^1} \left[ \frac{h}{T} \sum_{n=1}^{N-1} (1 + 4 \exp(-2\pi^2 M_2^2 r_1^{-2} h^{1-2\rho}))^{n-1} \right]. \quad (43) \end{aligned}$$

By (36), we have  $h \leq T$ . By the definition of  $\gamma_0$  in Lemma 5, we have that  $\gamma_0 = 2(1 - b)$  where  $b = 2M_3M_4a/M_2^2 = 2M_3M_4r_2h^{1/2}/M_2^2$ . Assumption (36) now implies that  $b \leq 1/2$  and  $\gamma_0^{-1} \leq 1$ . We write

$$c_\star = \sqrt{2}M_3M_2^{-1} \exp\left(\frac{r_2^2}{2M_2^2}\right) \exp\left(\frac{r_2^2(1 + TM_1)^2}{M_2^2}\right) T.$$

Let  $\xi(h) = 4 \exp(-c_1 h^{-c_2})$ , where  $c_1$  and  $c_2$  are positive constants with no dependence on  $h$ . We check that  $\xi$  satisfies the hypotheses of Lemma 1;  $h^{-\gamma}\xi(h)$  has a global maximum at  $h_* = (c_1 c_2 / \gamma)^{1/c_2}$ , and so we have  $\xi(h) \leq \epsilon h^\gamma$  for  $\epsilon = h_*^{-\gamma}\xi(h_*)$ , any choice of  $\gamma > 1$ , and all  $h > 0$ . With  $c_1 = 2\pi^2\gamma^2$  and  $c_2 = 2\rho - 1$ , we apply Lemma 1 to the term in square brackets on the right-hand side of (43). We conclude that

$$\frac{h}{T} \sum_{n=1}^{N-1} (1 + 4 \exp(-2\pi^2 M_2^2 r_1^{-2} h^{1-2\rho}))^{n-1} = 1 + o(h)$$

as  $h \rightarrow 0$  with  $N = T/h$ . By Lemma 2,  $k\|\hat{p}(\cdot, t_1)\|_{\ell^1} = 1 + o(k)$  as  $k \rightarrow 0$ . Putting everything together, we are left with (37).  $\square$

We are now in a position to combine our result with an earlier result from the literature [2] to establish the convergence of  $\hat{p}$  to  $p$ .

**Corollary 1.** *In addition to all of the hypotheses of Theorem 2, suppose that there exist constants  $\mathcal{F}_k, \mathcal{G}_k > 0$  such that*

$$\begin{aligned} \sup_{x \in \mathbb{R}} |f^{(k)}(x)| &\leq \mathcal{F}_k \\ \sup_{x \in \mathbb{R}} |g^{(k)}(x)| &\leq \mathcal{G}_k \end{aligned}$$

for all  $k \geq 0$ . Note that for  $k = 1$ , the first condition is redundant with (11a); for  $k = 0$  and  $k = 1$ , the second condition is redundant with (11b) and (11d). Then we have

$$\|p(\cdot, T) - \hat{p}(\cdot, T)\|_1 = O(h)$$

*Proof.* We have

$$\|p(\cdot, T) - \hat{p}(\cdot, T)\|_1 \leq \|p(\cdot, T) - \tilde{p}(\cdot, T)\|_1 + \|\tilde{p}(\cdot, T) - \hat{p}(\cdot, T)\|_1 \quad (44)$$

To handle the first term, we appeal to Corollary 2.1 from [2]. Our lower bound on  $g$  in (11b) corresponds to Bally and Talay's uniform ellipticity hypothesis "H1"; we may then apply Equations (27-28) from [2] to derive

$$|p(x, T) - \tilde{p}(x, T)| \leq h\mathcal{K}_1 \exp(-\mathcal{K}_2 x^2/T)$$

for constants  $\mathcal{K}_1, \mathcal{K}_2 > 0$  that do not depend on  $h$ . Therefore,

$$\|p(\cdot, T) - \tilde{p}(\cdot, T)\|_1 \leq h\mathcal{K}_1 \left(\frac{\pi T}{\mathcal{K}_2}\right)^{1/2} = O(h).$$

Returning to (44), by Theorem 2, the second term on the right-hand side goes to zero much faster than  $h$ , finishing the proof.  $\square$



## 6 Boundary Truncation

In practice, we do not evaluate (7) as it involves an infinite sum. In this section, we analyze a truncated version of the algorithm:

$$\mathring{p}(x, t_{n+1}) = k \sum_{j=-M}^M G(x, y_j) \mathring{p}(y_j, t_n) \quad (45)$$

This is the actual DTQ method used in practice. As in (8), we take  $\mathring{p}(x, t_1) = G(x, C)$  and use (45) starting with  $n = 1$ . Let us denote the error due to truncation by

$$r(x, t_{n+1}) = \hat{p}(x, t_{n+1}) - \mathring{p}(x, t_{n+1}) \quad (46)$$

By (8), we have  $r(x, t_1) \equiv 0$ . For  $n \geq 1$ , we have

$$r(x, t_{n+1}) = k \left( \sum_{|j|>M} G(x, y_j) \hat{p}(y_j, t_n) + \sum_{|j|\leq M} G(x, y_j) r(y_j, t_n) \right). \quad (47)$$

Based on the right-hand side, we see that it will be important to estimate the tail sum  $\sum_{|j|>M} \hat{p}(x_j, t_n)$ . We accomplish this using a Chernoff bound. To arrive at this bound, we construct a sequence of random variables  $\{Q_n\}_{n \geq 1}$ . We first define a normalization constant at time  $n$ :

$$K_n = \|\hat{p}(\cdot, t_n)\|_{\ell^1} = \sum_i \hat{p}(x_i, t_n). \quad (48)$$

By (16), we know that  $K_n < \infty$  for  $k > 0$  and  $h > 0$ . Let

$$q(x_i, t_n) = \frac{\hat{p}(x_i, t_n)}{K_n}, \quad (49)$$

so that  $\sum_i q(x_i, t_n) = 1$ . For each  $n$ , we postulate a random variable  $Q_n$  with state space  $\{k\mathbb{Z}\}$  and probability mass function  $q(\cdot, t_n)$ . In order to apply a Chernoff bound to  $Q_n$ , we must estimate its moment generating function.

**Lemma 7.** *Suppose  $f$  and  $g$  are admissible in the sense of Definition 2. Suppose  $k = h^\rho$  for some  $\rho > 1/2$ . Then there exists  $h_*$  such that for all  $h \in [0, h_*)$ , all  $s \in \mathbb{R}$ , and all  $n$  satisfying  $0 \leq n \leq (N-1)$ ,*

$$kE[e^{sQ_{n+1}}] < \frac{3}{2} \exp \left[ T \left( \frac{M_3^2 s^2}{2} + f(0)s \right) \right] \left( \frac{1}{2} + \exp(Cse^{M_1 T}) \right) < \infty.$$

*Proof.* We begin with our estimate of the moment generating function of  $Q_{n+1}$ . The calculation proceeds in two phases. The first phase is exact; note that in what follows we use the notation

$y_j = jk$ ,  $z_j = y_j + f(y_j)h$ , and  $g^2 = g^2(y_j)$ :

$$\begin{aligned}
E[e^{sQ_{n+1}}] &= \sum_{i=-\infty}^{\infty} e^{sx_i} q(x_i, t_{n+1}) \\
&= \frac{k}{K_{n+1}} \sum_i e^{sx_i} \sum_j \frac{1}{\sqrt{2\pi g^2 h}} \exp\left(-\frac{(x_i - z_j)^2}{2g^2 h}\right) \hat{p}(y_j, t_n) \\
&= \frac{k}{K_{n+1}} \sum_j \sum_i \frac{1}{\sqrt{2\pi g^2 h}} \exp\left(-\frac{x_i^2 - 2x_i z_j + z_j^2 - 2g^2 h s x_i}{2g^2 h}\right) \hat{p}(y_j, t_n) \\
&= \frac{1}{K_{n+1}} \sum_j \zeta_s(j) \exp\left(-\frac{z_j^2 - (z_j + g^2 h s)^2}{2g^2 h}\right) \hat{p}(y_j, t_n), \tag{50}
\end{aligned}$$

where

$$\zeta_s(j) = k \sum_i \frac{1}{\sqrt{2\pi g^2 h}} \exp\left(-\frac{(x_i - (z_j + g^2 h s))^2}{2g^2 h}\right).$$

It is at this point that we begin to estimate. Note that the summand is in fact a discrete Gaussian  $\phi(x_i)$ , as in (14), with  $\mu = z_j + g^2(y_j)hs$  and  $\sigma^2 = g^2(y_j)h$ . Hence we may apply the inequalities (15) and (11b) to write

$$\zeta_s(j) \leq 1 + 4 \exp\left(\frac{-2\pi^2 g^2(y_j)h}{k^2}\right) \leq 1 + 4 \exp\left(\frac{-2\pi^2 M_2^2 h}{k^2}\right). \tag{51}$$

Next, we turn our attention to the remaining exponential in (50). We use (11b), the mean value theorem, (11a), and the definition of  $z_j$  to obtain:

$$\begin{aligned}
\exp\left(-\frac{z_j^2 - (z_j + g^2 h s)^2}{2g^2 h}\right) &= \exp\left(z_j s + \frac{1}{2}g^2(y_j)h s^2\right) \\
&\leq e^{M_3^2 h s^2/2} \exp(y_j s + f(y_j)h s) \\
&\leq e^{M_3^2 h s^2/2} \exp(y_j s + f(0)h s + M_1 y_j h s) \\
&\leq e^{M_3^2 h s^2/2 + f(0)h s} \exp(y_j s(1 + M_1 h)) \tag{52}
\end{aligned}$$

Now we combine (50), (51), and (52). The result is

$$\begin{aligned}
E[e^{sQ_{n+1}}] &\leq \frac{K_n}{K_{n+1}} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2)) e^{M_3^2 h s^2/2 + f(0)h s} \\
&\quad \times \frac{1}{K_n} \sum_j \exp(y_j s(1 + M_1 h)) \hat{p}(y_j, t_n) \tag{53}
\end{aligned}$$

We recognize the expression on the second line as the moment generating function of  $Q_n$  evaluated at  $s' = s(1 + M_1 h)$ . Therefore,

$$\begin{aligned}
kE[e^{sQ_{n+1}}] &\leq \frac{K_n}{K_{n+1}} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2)) e^{M_3^2 h s^2/2 + f(0)h s} kE[e^{s(1+M_1 h)Q_n}] \\
&\leq \underbrace{\frac{K_1}{K_{n+1}} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^n}_{\zeta_1(h)} e^{T(M_3^2 s^2/2 + f(0)s)} \underbrace{kE[e^{s(1+M_1 h)Q_1}]}_{\zeta_2(h)}.
\end{aligned}$$

The main question now is what happens as  $h \rightarrow 0$  and  $N \rightarrow \infty$  such that  $hN = T$ . We assume that  $0 \leq n \leq (N - 1)$ . Because  $k = r_1 h^\rho$  for  $\rho > 1/2$ , we know by Lemma 3 that  $\zeta_1(h) \rightarrow 1$  as  $h \rightarrow 0$ . Hence there exists  $h_*^1$  such that  $h \in [0, h_*^1]$  ensures that  $|\zeta_1(h) - 1| < 1/2$ , i.e.,  $\zeta_1(h) < 3/2$ . Next, consider

$$\begin{aligned}\zeta_2(h) &= kE[e^{s(1+M_1h)^n Q_1}] \\ &= k \sum_{i=-\infty}^{\infty} e^{s(1+M_1h)^n x_i} \hat{p}(x_i, t_1) \\ &= k \sum_{i=-\infty}^{\infty} e^{s(1+M_1h)^n x_i} G(x_i, C) \\ &= \exp\left((C + f(C)h)s(1 + M_1h)^n + \frac{hg^2(C)s^2}{2}(1 + M_1h)^{2n}\right) k \sum_{i=-\infty}^{\infty} \phi(x_i),\end{aligned}$$

where  $\phi(x)$  is the Gaussian density defined in (14) with

$$\begin{aligned}\mu &= C + f(C)h + hg^2(C)s(1 + M_1h)^n \\ \sigma^2 &= hg^2(C)\end{aligned}$$

Now we apply Lemma 2 and  $n \leq (N - 1)$  to obtain

$$\begin{aligned}\zeta_2(h) &\leq \exp\left((C + f(C)h)s(1 + M_1h)^N + \frac{hg^2(C)s^2}{2}(1 + M_1h)^{2N}\right) \\ &\quad \times (1 + 4\exp(-2\pi^2 g^2(C)h/k^2)).\end{aligned}$$

As before,  $hk^{-2} = r_1^{-2}h^{1-2\rho} \rightarrow +\infty$  as  $h \rightarrow 0$ , and the term on the second line goes to 1 as  $h \rightarrow 0$ . Since  $\lim_{h \rightarrow 0^+} (1 + M_1h)^N = e^{M_1T}$ , we have

$$\lim_{h \rightarrow 0^+} \zeta_2(h) \leq \exp(Cse^{M_1T}).$$

Thus there exists  $h_*^2$  such that  $h \in [0, h_*^2]$  implies

$$|\zeta_2(h) - \exp(Cse^{M_1T})| \leq \frac{1}{2}.$$

Taking  $h_* = \min\{h_*^1, h_*^2\}$  finishes the proof.  $\square$

We can now give conditions under which  $r$ , defined in (46), converges to zero.

**Lemma 8.** *Suppose  $f$  and  $g$  are admissible in the sense of Definition 2. Suppose  $k = h^\rho$  for  $\rho > 1/2$ . For  $\varepsilon \geq 1$ , let*

$$M = \lceil (\varepsilon + \rho + 1)(-\log h)/k \rceil. \quad (54)$$

*Let  $h_*$  be defined as in Lemma 7. Then for  $h < h_*$ , we have  $k \sum_{|i| \leq M} |r(x_i, T)| = O(h)$ .*

*Proof.* We start with

$$|r(x_i, t_{n+1})| \leq k \sum_{|j|>M} G(x_i, y_j) \hat{p}(y_j, t_n) + k \sum_{|j|\leq M} G(x_i, y_j) |r(y_j, t_n)|.$$

Summing over  $i$ , we obtain

$$\sum_{|i|\leq M} |r(x_i, t_{n+1})| \leq k \sum_{|j|>M} \sum_{|i|\leq M} G(x_i, y_j) \hat{p}(y_j, t_n) + k \sum_{|j|\leq M} \sum_{|i|\leq M} G(x_i, y_j) |r(y_j, t_n)|.$$

Using (13) together with (11b), we have

$$\begin{aligned} \sum_{|i|\leq M} |r(x_i, t_{n+1})| &\leq (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2)) \sum_{|j|>M} \hat{p}(y_j, t_n) \\ &\quad + (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2)) \sum_{|j|\leq M} |r(y_j, t_n)|. \end{aligned} \quad (55)$$

This is of the form

$$r_{n+1} \leq \alpha \pi_n + \alpha r_n. \quad (56)$$

We derive from this the sequence of inequalities  $\alpha r_n \leq \alpha^2 \pi_{n-1} + \alpha^2 r_{n-1}, \dots, \alpha^{n-1} r_2 \leq \alpha^n \pi_1 + \alpha^n r_1$ .

Summing these together with (56), we derive  $r_{n+1} \leq \sum_{i=1}^n \alpha^i \pi_{n-i+1} + \alpha^n r_1$ . Applying this to (55)

and using  $r(\cdot, t_1) \equiv 0$ , we have

$$\sum_{|i|\leq M} |r(x_i, t_{n+1})| \leq \sum_{i=1}^n (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^i \sum_{|j|>M} \hat{p}(y_j, t_{n-i+1}). \quad (57)$$

Now we use (49) and the Chernoff bound to derive:

$$\begin{aligned} \sum_{|j|>M} \hat{p}(y_j, t_{n-i+1}) &= K_{n-i+1} \sum_{|j|>M} q(y_j, t_{n-i+1}) \\ &\leq K_{n-i+1} [P(Q_{n-i+1} \geq y_M) + P(Q_{n-i+1} \leq -y_M)] \\ &\leq K_{n-i+1} e^{-sy_M} (E[e^{sQ_{n-i+1}}] + E[e^{-sQ_{n-i+1}}]) \end{aligned}$$

We apply Lemma 7 to obtain

$$k \sum_{|j|>M} \hat{p}(y_j, t_{n-i+1}) \leq \frac{3}{2} K_{n-i+1} e^{-sy_M} \exp \left[ T \left( \frac{M_3^2 s^2}{2} + f(0)s \right) \right] (1 + 2 \cosh(Cse^{M_1 T})) \quad (58)$$

Applying this result to (57), we have

$$\begin{aligned} k \sum_{|i|\leq M} |r(x_i, t_{n+1})| &\leq \frac{3}{2} e^{-sy_M} \exp \left[ T \left( \frac{M_3^2 s^2}{2} + f(0)s \right) \right] (1 + 2 \cosh(Cse^{M_1 T})) \\ &\quad \times \sum_{i=1}^n (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^i K_{n-i+1}. \end{aligned}$$

By (48) and (16), we have

$$K_{n-i+1} \leq \|\hat{p}(\cdot, t_1)\|_{\ell^1} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{n-i}$$

Using this and  $n \leq N = T/h$ ,

$$\begin{aligned} k \sum_{|i| \leq M} |r(x_i, t_{n+1})| &\leq \frac{3}{2} e^{-sy_M} \exp \left[ T \left( \frac{M_3^2 s^2}{2} + f(0)s \right) \right] (1 + 2 \cosh(Cse^{M_1 T})) \\ &\quad \times \|\hat{p}(\cdot, t_1)\|_{\ell^1} \frac{T}{h} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{T/h}. \end{aligned} \quad (59)$$

Let  $s = 1$ . Note that  $\lim_{h \rightarrow 0} (1 + 4 \exp(-2\pi^2 M_2^2 h/k^2))^{T/h} = 1$  and  $\lim_{k \rightarrow 0} k \|\hat{p}(\cdot, t_1)\|_{\ell^1} = 1$ . Thanks to (54), we know that  $y_M \geq (\varepsilon + \rho + 1)(-\log h)$ . Putting things together, the right-hand side of (59) behaves like  $h^{\varepsilon+\rho+1} k^{-1} h^{-1} = h^\varepsilon = O(h)$  as desired.  $\square$

So long as  $M$  remains a positive integer, we can add/subtract a constant from (54) and still prove Lemma 8. What is important is how  $M$  scales as a function of  $h$ ; the logarithmic rate given in (54) is the rate at which we have to push  $M$  to  $+\infty$  so that we obtain  $O(h)$  convergence. If we push  $M$  to  $+\infty$  at a faster rate, e.g., by replacing  $(-\log h)$  with  $h^{-1}$ , then  $r$  will converge at a rate that is exponential in  $h$ .

Thus far we have considered convergence of  $r$  in a truncated and scaled version of the  $\ell^1$  norm. Convergence in  $L^1$  is an easy consequence.

**Lemma 9.** *Suppose  $f$  and  $g$  are admissible in the sense of Definition 2. Suppose  $k = h^\rho$  for  $\rho > 1/2$ . For  $\varepsilon \geq 1$ , let  $M$  be defined as in (54). Let  $h_*$  be defined as in Lemma 7. Then for  $h < h_*$ , we have  $\|r(\cdot, T)\|_1 = O(h)$ .*

*Proof.* Note that

$$|r(x, T)| \leq k \sum_{|j| > M} G(x, y_j) \hat{p}(y_j, t_{N-1}) + k \sum_{|j| \leq M} G(x, y_j) |r(y_j, t_{N-1})|.$$

This is similar to what we wrote above, except that the discrete variable  $x_i$  has been replaced by the continuous variable  $x$ . We now integrate both sides with respect to  $x$  to obtain

$$\|r(\cdot, T)\|_1 \leq k \sum_{|j| > M} \hat{p}(y_j, t_{N-1}) + k \sum_{|j| \leq M} |r(y_j, t_{N-1})|.$$

The second term is  $O(h)$  by Lemma 8. For the first term, we use (58) to write

$$k \sum_{|j| > M} \hat{p}(y_j, t_{N-1}) \leq \frac{3}{2} K_{N-1} e^{-y_M} \exp \left[ T \left( \frac{M_3^2}{2} + f(0) \right) \right] (1 + 2 \cosh(Ce^{M_1 T})). \quad (60)$$

Since  $\lim_{k \rightarrow 0} k K_{N-1} = 1$  and  $e^{-y_M} = O(h^{\varepsilon+\rho+1})$ , the right-hand side of (60) behaves like  $h^{\varepsilon+1} = O(h^2)$ .  $\square$

It is now immediately clear that, under certain conditions, we have established  $O(h)$  convergence of  $\hat{p}$  to the true density  $p$  in the  $L^1$  norm.

**Corollary 2.** *Suppose that all of the hypotheses of Corollary 1 and Lemma 9 are satisfied. Then, combining these results, we have  $\|p(\cdot, T) - \hat{p}(\cdot, T)\|_1 = O(h)$ .*

## 7 Numerical Experiments

In this section, we use R/C++ implementations of the DTQ method to study its empirical convergence behavior, and also to compare against a numerical solver for (2), the Fokker-Planck or Kolmogorov equation. All codes described in this section, together with instructions on how to reproduce Figures 1 and 2, are available at the following URL:

<https://github.com/hbhat4000/sdeinference/tree/master/DTQpaper>

We caution the reader that, in the present work, we do not deal with all important implementation issues. Here we are primarily concerned with demonstrating properties of the DTQ method. This can be done quite well even with the assumptions on the initial condition and domain sizes given below. Relaxing these assumptions poses no conceptual difficulties, but may require changes to technical details in the codes linked above.

### 7.1 Convergence

First, we compare empirical and theoretical convergence behavior. We verify that under the conditions given by Theorem 2, we do observe convergence in practice. We also show numerical evidence that such convergence takes place when one or more of the hypotheses do not hold.

All of the SDE we consider are equations for a scalar unknown  $X_t$ . We describe here the way in which we conduct numerical tests for each SDE. We begin with the initial condition  $X_0 = 0$  and solve forward in time until  $T = 1$ . That is, we apply the DTQ method to compute  $\hat{p}(x, 1)$ . We use the following values of the temporal step  $h$ :

$$\{0.5, 0.2, 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001\}. \quad (61)$$

For  $h \geq 0.01$ , we find that an implementation of the DTQ method written completely in R is able to run in a reasonable amount of time. For  $h = 0.005$  and below, we use an implementation where computationally intensive parts of the code are written in C++; this code is glued to our R code using the Rcpp and RcppArmadillo packages [12, 11, 13, 28].

The remaining algorithm parameters are set in the following way:

$$k = h^{3/4} \quad (62a)$$

$$\begin{cases} \text{Examples 1,2,4,5,6} & M = \lceil \pi/k^2 \rceil \\ \text{Example 3} & M = \lceil \pi/(2k) - 2 \rceil. \end{cases} \quad (62b)$$

$$x_j = jk, \text{ for } -M \leq j \leq M. \quad (62c)$$

For each value of  $h$ , we compare  $\hat{p}(x, T)$  computed using the DTQ method against the exact solution  $p(x, T)$ . Let  $F(y, T) = \int_{x=-\infty}^{x=y} p(x, T) dx$  denote the cumulative distribution function associated with the density  $p$ . Each comparison is carried out using the following three norms:

$$\|p(\cdot, T) - \hat{p}(\cdot, T)\|_1 \approx k \sum_{j=-M}^{j=M} |p(jk, T) - \hat{p}(jk, T)| \quad (63a)$$

$$\|p(\cdot, T) - \hat{p}(\cdot, T)\|_\infty \approx \sup_{|j| \leq M} |p(jk, T) - \hat{p}(jk, T)| \quad (63b)$$

$$\|F(\cdot, T) - \hat{F}(\cdot, T)\|_\infty \approx \sup_{|j| \leq M} |F(jk, T) - \hat{F}(jk, T)| \quad (63c)$$

For our tests, we consider six SDE, all for a scalar unknown  $X_t$ :

$$\text{Example 1: } \begin{cases} dX_t = -X_t dt + dW_t \\ p(x, t) = \frac{\exp(-x^2/(1 - \exp(-2t)))}{\sqrt{(\pi(1 - \exp(-2t)))}} \end{cases} \quad (64a)$$

$$\text{Example 2: } \begin{cases} dX_t = -\frac{1}{2} \tanh X_t \operatorname{sech}^2 X_t dt + \operatorname{sech} X_t dW_t \\ p(x, t) = (2\pi t)^{-1/2} (\cosh x) \exp(-\sinh^2 x/(2t)) \end{cases} \quad (64b)$$

$$\text{Example 3: } \begin{cases} dX_t = -(\sin X_t \cos^3 X_t) dt + (\cos^2 X_t) dW_t \\ p(x, t) = (2\pi t)^{-1/2} (\sec^2 x) \exp(-\tan^2 x/(2t)) \end{cases} \quad (64c)$$

$$\text{Example 4: } \begin{cases} dX_t = \left(\frac{1}{2} X_t + \sqrt{1 + X_t^2}\right) dt + \sqrt{1 + X_t^2} dW_t \\ p(x, t) = (2\pi(1 + x^2))^{-1/2} \exp(-(\sinh^{-1} x - t)^2/2) \end{cases} \quad (64d)$$

$$\text{Example 5: } \begin{cases} dX_t = \frac{1}{2} X_t dt + \sqrt{1 + X_t^2} dW_t \\ p(x, t) = (2\pi t(1 + x^2))^{-1/2} \exp(-(\sinh^{-1} x)^2/(2t)) \end{cases} \quad (64e)$$

$$\text{Example 6: } \begin{cases} dX_t = \left(-\sqrt{1 + X_t^2} \sinh^{-1} X_t + \frac{1}{2} X_t\right) dt + \sqrt{1 + X_t^2} dW_t \\ p(x, t) = \frac{\exp(-(\sinh^{-1} x)^2/(1 - \exp(-2t)))}{\sqrt{(\pi(1 - \exp(-2t))(1 + x^2))}} \end{cases} \quad (64f)$$

Note that for each example, we have supplied an exact solution in the form of a probability density function  $p(x, t)$ . For each example, we compare the DTQ density with  $p(x, T = 1)$ .

Figure 1 shows the convergence results for all six examples. The overall impression we gain from the plots is that the practical  $L^1$  error between the DTQ and exact density functions scales like  $h$ . As we now explain, this first-order convergence is displayed under a variety of conditions.

Example 1 features drift and diffusion coefficients that clearly satisfy the hypotheses of our convergence theory. In this case, the computational results confirm the theory.

In Example 2, the drift and diffusion coefficients satisfy all but one of the hypotheses. Specifically, because  $\operatorname{sech} x \rightarrow 0$  as  $|x| \rightarrow \infty$ , the diffusion coefficient is not bounded away from zero. However, as a matter of numerical practice, on any truncated domain of the form (62), the diffusion coefficient never equals zero. We can say, then, that on the computational domain, the diffusion coefficient does have a global lower bound that is greater than zero. The computational results display first-order convergence.

Example 3 is similar to Example 2 in that all but one of the hypotheses are satisfied. Again, it is the diffusion coefficient  $\cos^2 y$  that is not bounded away from zero. However, either an analysis of the original SDE or inspection of the exact solution reveals that the density will only be supported on the interval  $(-\pi/2, \pi/2)$ . For this SDE, we set  $M = \lceil \pi/(2k) - 2 \rceil$  as in (62b), retaining (62a) and (62c). This way, the spatial grid covers the interior of  $(-\pi/2, \pi/2)$  and the diffusion coefficient never reaches zero. Again, the computational results show that the  $L^1$  error scales like  $h$ .

Moving to Examples 4 and 5, we now have instances where the diffusion coefficient is bounded from below by 1 but is unbounded above. All other hypotheses of our convergence theory are satisfied. The empirical convergence rates for both examples match what we expect from theory.

Reexamining the situation with slightly more depth, what we find from our proofs is that (24) is the only place where the upper bound on the diffusion coefficient is used. However, for the particular case of the diffusion coefficient  $g(x) = (1 + x^2)^{1/2}$  used in Examples 4 and 5, we have

that

$$|\Im(g(y + i\epsilon)g'(y + i\epsilon))| = |\Im(y + i\epsilon)| \leq d,$$

meaning that we can substitute  $d$  for  $M_3M_4$  and the convergence proof follows. This is an example of how, for specific SDE that do not satisfy the hypotheses of the general theorem, we may yet be able to prove convergence of the DTQ method.

Finally, we come to Example 6. Now we have that derivative of the drift coefficient is unbounded *and* that the diffusion coefficient is unbounded above. Still, the results in the convergence plot agree with the overall first-order convergence rate implied by theory.

For the SDE in Example 6, even if we are able to patch our proof to prove that  $\hat{p}$  converges to  $\tilde{p}$ , we can no longer apply the result of Bally and Talay [2] to guarantee convergence of  $\tilde{p}$  to  $p$ . Overall, we take the numerical results for Example 6 as evidence that  $\tilde{p}$  must converge to  $p$  under more general conditions than have been established in the literature.

## 7.2 Comparison with Fokker-Planck

Now we turn to a comparison of the DTQ method with a classical approach, that of numerically solving the Fokker-Planck or Kolmogorov PDE (2). In what follows, we use subscripts to denote partial derivatives, so that (2) is written

$$p_t + (f(x)p(x, t))_x = \frac{1}{2} (g^2(x)p(x, t))_{xx}. \quad (65)$$

To solve this equation, we employ a standard finite difference method. To resolve the singular initial condition  $p(x, 0) = \delta(x)$ , we use a standard subtraction idea: we set  $p = u + v$ , where  $u$  solves

$$u_t = \frac{1}{2} \kappa u_{xx} \quad (66a)$$

$$u(x, 0) = \delta(x), \quad (66b)$$

while  $v$  solves

$$v_t + (f(x)v(x, t))_x = \frac{1}{2} (g^2(x)v(x, t))_{xx} + \underbrace{\frac{1}{2} [(g^2(x) - \kappa) u(x, t)]_{xx} - [f(x)u(x, t)]_x}_{F(x, t)} \quad (67a)$$

$$v(x, 0) = 0. \quad (67b)$$

The point is that (66) can be solved analytically, i.e., for  $t > 0$ ,

$$u(x, t) = \frac{1}{\sqrt{2\pi\kappa t}} \exp\left(-\frac{x^2}{2\kappa t}\right). \quad (68)$$

Here  $\kappa > 0$  is a parameter that we are free to set. In our own tests, we use  $\kappa = 1$ . Since (68) is known, we substitute it into the final two terms on the right-hand side of (67a)—this yields a known forcing term  $F(x, t)$ . We then employ the following numerical scheme to solve (67) for  $v(x, t)$ :

- We discretize  $v(x, t)$  on fixed spatial and temporal grids with respective spacings  $k$  and  $h$ . Let  $V_j^n$  denote our numerical approximation to  $v(jk, nh)$ . Here  $0 \leq n \leq N$  with  $Nh = T > 0$ , the final time. We also have that  $-M \leq j \leq M$ . Implicitly, we assume that  $v(x, t) = 0$  for  $|x| > Mk$ .



- We use a first-order approximation to  $v_t$ :  $v_t(x, t) \approx (V_j^{n+1} - V_j^n)/h$ .
- We treat the drift term explicitly:

$$(f(x)v(x, t))_x \approx (f((j+1)k)V_{j+1}^n - f((j-1)k)V_{j-1}^n)/(2k).$$

- We treat the diffusion term implicitly:

$$\frac{1}{2} (g^2(x)v(x, t))_{xx} \approx \frac{1}{2k^2} (g^2((j-1)k)V_{j-1}^{n+1} - 2g^2(jk)V_j^{n+1} + g^2((j+1)k)V_{j+1}^{n+1}).$$

Let  $\mathbf{V}^n$  be a vector of length  $2M+1$  whose  $j$ -th entry is  $V_j^n$ . Then, combining approximations, we obtain the matrix-vector system

$$A\mathbf{V}^{n+1} = B\mathbf{V}^n + \mathbf{F}^n \quad (69)$$

with tridiagonal matrices  $A$  and  $B$  given by

$$A = \begin{bmatrix} 1 + \frac{h}{k^2}g_{-M}^2 & -\frac{h}{2k^2}g_{-M+1}^2 & & & \\ -\frac{h}{2k^2}g_{-M}^2 & 1 + \frac{h}{k^2}g_{-M+1}^2 & -\frac{h}{2k^2}g_{-M+2}^2 & & \\ & -\frac{h}{2k^2}g_{-M+1}^2 & 1 + \frac{h}{k^2}g_{-M+2}^2 & -\frac{h}{2k^2}g_{-M+3}^2 & \\ & & \ddots & \ddots & \ddots \\ & & & -\frac{h}{2k^2}g_{M-2}^2 & 1 + \frac{h}{k^2}g_{M-1}^2 & -\frac{h}{2k^2}g_M^2 \\ & & & & -\frac{h}{2k^2}g_{M-1}^2 & 1 + \frac{h}{k^2}g_M^2 \end{bmatrix} \quad (70)$$

and

$$B = \begin{bmatrix} 1 & -\frac{h}{2k}f_{-M+1} & & & \\ \frac{h}{2k}f_{-M} & 1 & -\frac{h}{2k}f_{-M+2} & & \\ & \frac{h}{2k}f_{-M+1} & 1 & -\frac{h}{2k}f_{-M+3} & \\ & & \ddots & \ddots & \ddots \\ & & & \frac{h}{2k}f_{M-2} & 1 & -\frac{h}{2k}f_M \\ & & & & \frac{h}{2k}f_{M-1} & 1 \end{bmatrix}. \quad (71)$$

We also define  $\mathbf{F}^n$  in (69) by discretizing  $F(x, t)$  in (67a). Specifically, for  $-M \leq j \leq M$ , we define the  $j$ -th component of  $\mathbf{F}^n$  by

$$F_j^n = \frac{h}{2k^2} [g^2((j-1)k)u((j-1)k, nh) - 2g^2(jk)u(jk, nh) + g^2((j+1)k)u((j+1)k, nh)] \\ - \frac{h}{2k} [f((j+1)k)u((j+1)k, nh) - f((j-1)k)u((j-1)k, nh)]. \quad (72)$$

To solve for  $\mathbf{V}^{n+1}$  given  $\mathbf{V}^n$ , we rewrite (69) as

$$\mathbf{V}^{n+1} = A^{-1}B\mathbf{V}^n + A^{-1}\mathbf{F}^n. \quad (73)$$

Let  $\mathbf{U}^N$  denote the vector obtained by evaluating  $u(jk, T)$  for  $-M \leq j \leq M$ . Let  $p_{\text{FP}}(\mathbf{x}, T)$  denote the vector whose  $j$ -th component is  $p_{\text{FP}}(x_j, T)$ , the approximation of  $p(x_j, T)$  obtained by solving the Fokker-Planck equation numerically. With these definitions, our algorithm for computing  $p_{\text{FP}}$  is easily stated: we start with  $\mathbf{V}^0 = \mathbf{0}$ , iterate (73)  $N$  times to compute  $\mathbf{V}^N$ , and then compute

$$p_{\text{FP}}(\mathbf{x}, T) = \mathbf{U}^N + \mathbf{V}^N.$$

Note that in our implementation of the Fokker-Planck method, the matrices  $A$  and  $B$  defined by (70) and (71) are implemented as sparse tridiagonal matrices. When we use (73) to solve for  $\mathbf{V}^{n+1}$ , we use sparse numerical linear algebra to compute both  $A^{-1}B$  and  $A^{-1}\mathbf{F}^n$ . In particular,  $A^{-1}B$  is precomputed before we loop from  $n = 0$  to  $n = N - 1$ .

We are now in a position to compare the DTQ and Fokker-Planck methods. For this comparison, we exclusively use the drift and diffusion functions from Example 1 in (64). As described above, among the examples in (64), Example 1 is the only one that satisfies all of the hypotheses of our DTQ convergence theory.

As mentioned in Section 6, when we implement the DTQ method in practice, we start with (45)—with  $x$  discretized on the same spatial grid as  $y$ , i.e.,

$$\dot{p}(x_i, t_{n+1}) = k \sum_{j=-M}^M G(x_i, y_j) \dot{p}(y_j, t_n) \quad (74)$$

For fixed  $n$ , as  $j$  varies from  $-M$  to  $M$ , the elements  $\dot{p}(y_j, t_n)$  form a  $(2M + 1)$ -dimensional vector that we denote  $\mathbf{p}^n$ . With this notation, (74) can be written

$$\mathbf{p}^{n+1} = \mathcal{A}\mathbf{p}^n, \quad (75)$$

where  $\mathcal{A}$  is the  $(2M + 1) \times (2M + 1)$  matrix whose  $(i, j)$ -th element is  $kG(x_i, y_j)$ . In our experience, *the most computationally expensive part of the DTQ method is the assembly of  $\mathcal{A}$* . For the tests presented in this subsection, we have implemented three different methods to compute  $\mathcal{A}$ :

1. **DTQ-Naïve.** Here we assemble  $\mathcal{A}$  using dense matrix methods in R. The main advantage of this approach is ease of implementation; the code to compute  $\mathcal{A}$  is only 4 lines long. Incidentally, the convergence tests in the first part of this section use the DTQ-Naïve method for  $h \geq 0.01$ .
2. **DTQ-CPP.** Implicitly, the DTQ-Naïve method forces R to loop over the entries of  $\mathcal{A}$  serially. In the DTQ-CPP method, we use Rcpp together with OpenMP directives to compute and fill in the entries of  $\mathcal{A}$  in parallel. In practice, we run this code on a machine with 12 cores, setting the number of OpenMP threads to 12.
3. **DTQ-Sparse.** Here we take advantage of the structure of  $\mathcal{A}$ . Specifically, we have

$$\mathcal{A}_{ij} = kG(x_i, y_j) = \frac{k}{\sqrt{2\pi g^2(y_j)h}} \exp\left(-\frac{(x_i - y_j - f(y_j)h)^2}{2g^2(y_j)h}\right).$$

Let us set  $i = j + i'$ . Then we have

$$\mathcal{A}_{j+i',j} = \frac{k}{\sqrt{2\pi g^2(y_j)h}} \exp\left(-\frac{(i'k - f(y_j)h)^2}{2g^2(y_j)h}\right). \quad (76)$$

We think of  $i'$  as indexing the sub-/super-diagonals of  $\mathcal{A}$ . For each fixed  $i' = 0, 1, 2, \dots$  we evaluate (76) over all  $j$  to obtain the  $i'$ -th subdiagonal of  $\mathcal{A}$ . For  $h$  small, as  $i'$  increases, we observe that the entire subdiagonal decays rapidly. In our implementation, we compute subdiagonals until the 1-norm of the subdiagonal drops below  $2.2 \times 10^{-16}$  (machine precision in R) multiplied by the 1-norm of the main  $i' = 0$  diagonal of  $\mathcal{A}$ . We then compute the same number of superdiagonals as subdiagonals. The final  $\mathcal{A}$  matrix is assembled as a sparse matrix using the CRAN Matrix package [3].

Given the tridiagonal structure of both  $A$  and  $B$  in the Fokker-Planck method, we do not believe any reasonable modern implementation would use dense matrices. Similarly, while DTQ-Naïve requires minimal programming effort, a reasonable implementation would look much more like DTQ-CPP or DTQ-Sparse. None of the DTQ methods require more programming effort to implement than the Fokker-Planck method.

**Results for  $O(h^{3/4})$  Domain Scaling.** For each  $h$  in (61) that satisfies  $h \geq 0.01$ , we use all three DTQ methods and the Fokker-Planck method to generate numerical approximations of the density function at the final time  $T = 1$ . For our first set of comparisons, parameters such as  $k$  and  $M$  are set via (62). In particular, the computational domain is  $[-y_M, y_M]$  where  $y_M = Mk \propto h^{-3/4}$ . We compute the  $L^1$  errors between each numerical solution and the exact solution  $p(x, T)$ . We also record the wall clock time (in seconds) required to compute the solution using each method. Each measurement is repeated 100 times; we report average results.

In the left panel of Figure 2, we have plotted (on log-scaled axes) wall clock time as a function of  $L^1$  error for each of the four methods. We see that if one can tolerate a relatively large  $L^1$  error, then the fastest method is the DTQ-Naïve method (green); for  $L^1$  errors less than 0.03, the fastest method is the DTQ-Sparse method (purple). The Fokker-Planck method is often the slowest of the four methods. For an error of 0.003, the DTQ-Sparse method is approximately 100 times faster than the Fokker-Planck method.

**Results for  $O(\log h^{-1})$  Domain Scaling.** For our second set of comparisons, we have changed the way that  $y_M$  (effectively, the size of the computational domain) scales with  $h$ . We retain  $k = h^{3/4}$  but now set  $y_M = (2 + 3/4)(-\log h) \propto (-\log h)$  in accordance with (54). The spatial grid, for all four methods, is now given by  $x_j = -y_M + (j + M)k$  for  $-M \leq j \leq M$  with  $M = \lfloor y_M/k \rfloor$ . In all other respects, we make no changes and rerun the test described above for all four methods.

In the right panel of Figure 2, we have plotted (on log-scaled axes) wall clock time as a function of  $L^1$  error for each of the four methods. Once again, we find that the DTQ-Naïve and DTQ-Sparse methods are the fastest for, respectively, large and small error values. For an error of 0.003, the DTQ-Sparse method is approximately  $10^{3/4} \approx 5.62$  times faster than the Fokker-Planck method.

## 8 Conclusion and Future Directions

In this paper, we have established fundamental properties of the DTQ method, including theoretical and empirical convergence results. The present research motivates four main questions that we seek to answer in future work. Before getting into these questions, let us make three concluding remarks regarding our results.

First, until now we have not mentioned that the DTQ method features two properties that are not always easy to establish for numerical methods for the Fokker-Planck equation (2): (i) the DTQ method automatically preserves the nonnegativity of the computed density  $\hat{p}$ , and (ii) the DTQ density  $\hat{p}$  has a normalization constant that can be estimated for finite  $h, k > 0$ . In practice, we find that  $\hat{p}$  is very close to being correctly normalized.

Second,  $p(x, T)$  and  $\tilde{p}(x, t_N)$  correspond to, respectively, the random variables  $X_T$  and  $x_N$ .

Convergence in  $L^1$  of  $\tilde{p}$  to  $p$  is equivalent to convergence in total variation of  $x_N$  to  $X_T$ . Note that

$$\int_{x=-\infty}^{\infty} \hat{p}(x, t_{n+1}) dx = k \sum_{j=-\infty}^{\infty} \hat{p}(y_j, t_n) = kK_n, \quad (77)$$

implying that  $\hat{q}(x, t_{n+1}) = \hat{p}(x, t_{n+1})/(kK_n)$  is the density function of a continuous random variable  $y_n$ . An easy consequence of our results is that  $\hat{q}$  converges to  $\tilde{p}$  in  $L^1$ , implying convergence of  $y_N$  to  $x_N$  in total variation.

Third, if we trace back the crux of our convergence proof, a key step is estimating the  $L^1$  error of  $\tau$  starting from the trapezoidal rule error estimate (41). To do this, it was essential that we have an estimate of  $\mathcal{N}$  that is an  $L^1$  function of  $x$ . It was to obtain such an estimate that we put our efforts into Lemma 5. We have tried to replicate this analysis using more conventional error estimates for the trapezoidal rule—estimates that require less regularity of the integrand than we have assumed. Thus far, these other attempts have failed because they do not yield an upper bound on  $\tau$  that is itself an  $L^1$  function of  $x$ . The approach in the present work is the only one that we have gotten to work.

As mentioned above, there are four main questions that the present work motivates. These questions are the subject of future work:

1. When we derived the DTQ method, we used three approximations: (i) an Euler-Maruyama approximation of the original SDE, (ii) a trapezoidal quadrature rule, and (iii) a finite-dimensionalization of  $\tilde{p}$  that consists of sampling the function on a truncated grid. The first question to ask is: what happens to the DTQ method if we improve upon these initial approximations?

Regarding (ii), we can say that we have written a test code in which we use Gauss-Hermite quadrature instead of the trapezoidal rule. This does not yield better convergence. Given the exponential convergence of  $\hat{p}$  to  $\tilde{p}$  established here, this should not be a surprise.

Regarding (iii), rather than sampling the function  $\tilde{p}(x, t_n)$  on a discrete grid, we could have instead chosen to represent  $\tilde{p}(x, t_n)$  as a linear combination of functions—for instance, a linear combination of Gaussian densities, where each density is centered at a grid point  $x_j$ . In a collocation scheme, we would then insert these approximations of  $\tilde{p}$  into (6) and enforce equality at a finite number of points. We have tried this as well in a test code. While such a scheme does not yield better numerical behavior, it may be easier to analyze.

Approximation (i) is the one that would most easily yield major improvements. In the DTQ derivation, we can easily replace the Euler-Maruyama method with a higher-order method. The only change is to then replace the Gaussian kernel  $G$  with a different conditional density function. With this new  $G$ , the evolution equation (45) for  $\hat{p}$  remains the same. Preliminary results with the weak trapezoidal method [1] indicate that, in this way, we can obtain a version of the DTQ method that features  $O(h^2)$  convergence of  $\hat{p}$  to  $p$ .

2. Can we patch the DTQ method to handle diffusion functions  $g$  that equal zero at, say, a finite number of discrete points in the computational domain? We believe there should be some way of doing this by subtracting out singularities of  $G$  inside the Chapman-Kolmogorov equation (7).

3. Can we derive DTQ-like methods for stochastic differential equations driven by stochastic processes other than the Wiener process? In ongoing work, we are studying how to derive such methods to solve for the density in the case when we replace  $dW_t$  by a process whose increments follow a Lévy  $\alpha$ -stable distribution. For such an SDE, current methods for computing the density involve numerical solution of a fractional Fokker-Planck equation. We expect DTQ-like methods to be highly competitive for such problems.
4. How can we further apply the DTQ method to problems of statistical inference? In a typical inference problem, we seek to use data to infer parameters in the drift and diffusion functions. In preliminary work, we have shown how the DTQ method can be used to efficiently compute two quantities that are important for inference: the likelihood function and its gradient with respect to the parameters [7, 8]. Further improvements to and generalizations of the DTQ method, as described above, will yield improved inference algorithms.

## Acknowledgments

H.S.B. and R.W.M.A.M. gratefully acknowledge support for this work from UC Merced, through a UC Merced Committee on Research grant, USAP summer fellowships, and Applied Mathematics Graduate Group travel grants.

## References

- [1] DAVID F. ANDERSON AND JONATHAN C. MATTINGLY, *A weak trapezoidal method for a class of stochastic differential equations*, 9 (2011), pp. 301–318.
- [2] VLAD BALLY AND DENIS TALAY, *The law of the Euler scheme for stochastic differential equations. II. Convergence rate of the density*, Monte Carlo Methods and Applications, 2 (1996), pp. 93–128.
- [3] DOUGLAS BATES AND MARTIN MAECHLER, *Matrix: Sparse and Dense Matrix Classes and Methods*, 2016. R package version 1.2-4.
- [4] HARISH S. BHAT, *Algorithms for linear stochastic delay differential equations*, in Topics in Statistical Simulation, V. B. Melas, Stefania Mignani, Paola Monari, and Luigi Salmaso, eds., vol. 114 of Springer Proceedings in Mathematics & Statistics, Springer New York, 2014, pp. 57–65.
- [5] HARISH S. BHAT AND NITESH KUMAR, *Spectral solution of delayed random walks*, Phys. Rev. E, 86 (2012), p. 045701.
- [6] HARISH S. BHAT AND R. W. M. A. MADUSHANI, *Computing the density function for a nonlinear stochastic delay system*, IFAC-PapersOnLine, 48 (2015), pp. 316–321. 12th IFAC Workshop on Time Delay Systems (TDS 2015), Ann Arbor, Michigan, USA, 28-30 June 2015.
- [7] H. S. BHAT AND R. W. M. A. MADUSHANI, *Nonparametric adjoint-based inference for stochastic differential equations*, in Proceedings of the 3rd IEEE International Conference on Data Science and Advanced Analytics, Special Session on Statistical Learning for Data Science, 2016. Accepted for publication.

- [8] H. S. BHAT, R. W. M. A. MADUSHANI, AND S. RAWAT, *Scalable SDE filtering and inference with Apache Spark*, Journal of Machine Learning Research: Workshop and Conference Proceedings, 53 (2016). In press.
- [9] YUZHAI CAI, *Convergence theory of a numerical method for solving the Chapman-Kolmogorov equation*, SIAM Journal on Numerical Analysis, 40 (2003), pp. 2337–2351.
- [10] T. CANOR AND V. DENOËL, *Transient Fokker-Planck-Kolmogorov equation solved with smoothed particle hydrodynamics method*, International Journal for Numerical Methods in Engineering, 94 (2013), pp. 535–553.
- [11] DIRK EDDERBUETTEL, *Seamless R and C++ Integration with Rcpp*, Springer, New York, 2013.
- [12] DIRK EDDERBUETTEL AND ROMAIN FRANÇOIS, *Rcpp: Seamless R and C++ integration*, Journal of Statistical Software, 40 (2011), pp. 1–18.
- [13] DIRK EDDERBUETTEL AND CONRAD SANDERSON, *RcppArmadillo: Accelerating R with high-performance C++ linear algebra*, Computational Statistics and Data Analysis, 71 (2014), pp. 1054–1063.
- [14] C. FUCHS, *Inference for Diffusion Processes: With Applications in Life Sciences*, Springer, Berlin, 2013.
- [15] M. B. GILES, T. NAGAPETIAN, AND K. RITTER, *Multilevel Monte Carlo approximation of distribution functions and densities*, SIAM/ASA J. Uncertainty Quantification, 3 (2015), pp. 267–295.
- [16] Y. HU AND S. WATANABE, *Donsker delta functions and approximations of heat kernels by the time discretization method*, J. Math. Kyoto Univ., 36 (1996), pp. 494–518.
- [17] A. S. HURN, J. I. JEISMAN, AND K. A. LINDSAY, *Seeing the wood for the trees: A critical evaluation of methods to estimate the parameters of stochastic differential equations*, Journal of Financial Econometrics, 5 (2007), pp. 390–455.
- [18] A. KOHATSU-HIGA, *High order Ito-Taylor approximations to heat kernels*, J. Math. Kyoto Univ., 37 (1997), pp. 129–150.
- [19] S. C. KOU, B. P. OLDING, M. LYSY, AND J. S. LIU, *A multiresolution method for parameter estimation of diffusion processes*, Journal of the American Statistical Association, 107 (2012), pp. 1558–1574.
- [20] HAROLD J. KUSHNER, *On the weak convergence of interpolated Markov chains to a diffusion*, Ann. Probability, 2 (1974), pp. 40–50.
- [21] JOHN LUND AND KENNETH L. BOWERS, *Sinc methods for quadrature and differential equations*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.
- [22] X. LUO AND S. S.-T. YAU, *Hermite spectral method to 1D forward Kolmogorov equation and its application to nonlinear filtering problems*, IEEE Transactions on Automatic Control, 58 (2013), pp. 2495–2507.

- [23] G. N. MILSTEIN, J. G. M. SCHOENMAKERS, AND V. SPOKOINY, *Transition density estimation for stochastic differential equations via forward-reverse representations*, Bernoulli, 10 (2004), pp. 281–312.
- [24] M. DI PAOLA AND A. SOFI, *Approximate solution of the Fokker-Planck-Kolmogorov equation*, Probabilistic Engineering Mechanics, 17 (2002), pp. 369–384.
- [25] A. R. PEDERSEN, *A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations*, Scandinavian Journal of Statistics, 22 (1995), pp. 55–71.
- [26] L. PICHLER, A. MASUD, AND L. A. BERGMAN, *Numerical solution of the Fokker-Planck equation by finite differences and finite element methods—a comparative study*, in Computational Methods in Stochastic Dynamics, vol. 2, Springer, 2013, pp. 69–85.
- [27] L. C. G. ROGERS, *Smooth transition densities for one-dimensional diffusions*, Bull. London Math. Soc., 17 (1985), pp. 157–161.
- [28] CONRAD SANDERSON AND RYAN CURTIN, *Armadillo: a template-based C++ library for linear algebra*, Journal of Open Source Software, 1 (2016), p. 26.
- [29] P. SANTA-CLARA, *Simulated likelihood estimation of diffusions with an application to the short term interest rate*, Tech. Report 12-97, Anderson School of Management, UCLA, Los Angeles, California, 1997.
- [30] S. J. SHEATHER AND M. C. JONES, *A reliable data-based bandwidth selection method for kernel density estimation*, J. Roy. Statist. Soc. Ser. B, 53 (1991), pp. 683–690.
- [31] FRANK STENGER, *Numerical Methods Based on Sinc and Analytic Functions*, Springer Series in Computational Mathematics, Springer, New York, 2012.
- [32] LLOYD N. TREFETHEN AND J. A. C. WEIDEMAN, *The exponentially convergent trapezoidal rule*, SIAM Review, 56 (2014), pp. 385–458.

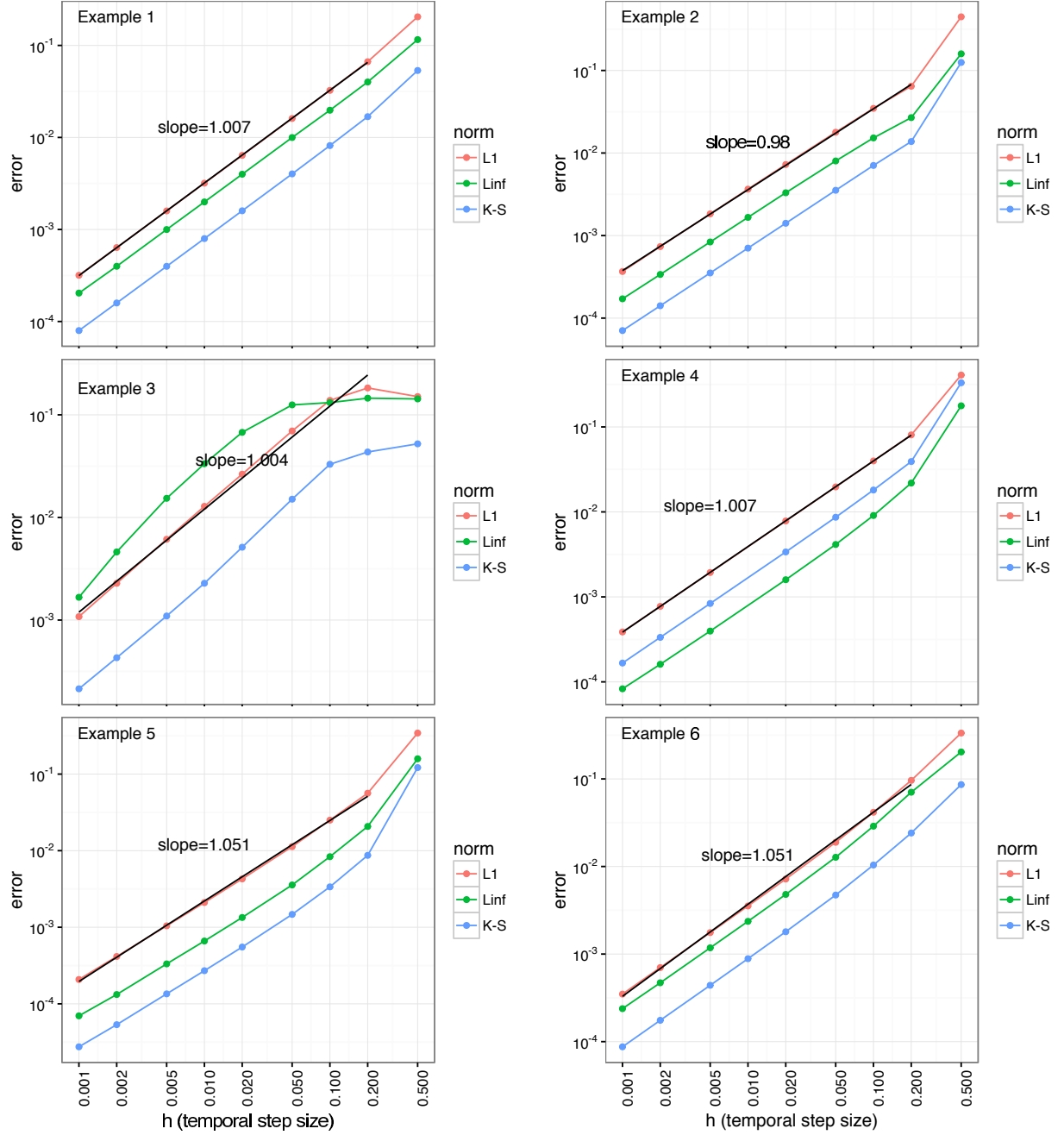


Figure 1: For each of the six examples in (64), we test the DTQ method's convergence. For each example, we plot errors between DTQ and exact solutions on log-scaled axes as a function of  $h$ , the temporal step size; all other parameters are given by (62). We compute errors in each of the three norms given by (63). The horizontal axes (labels and tick mark locations) are the same for all plots and correspond to the  $h$  values in (61). Least-squares fits to the  $L^1$  error data are indicated by black lines and corresponding slope values. For all examples, we observe first-order convergence, consistent with our  $O(h)$  theoretical result.



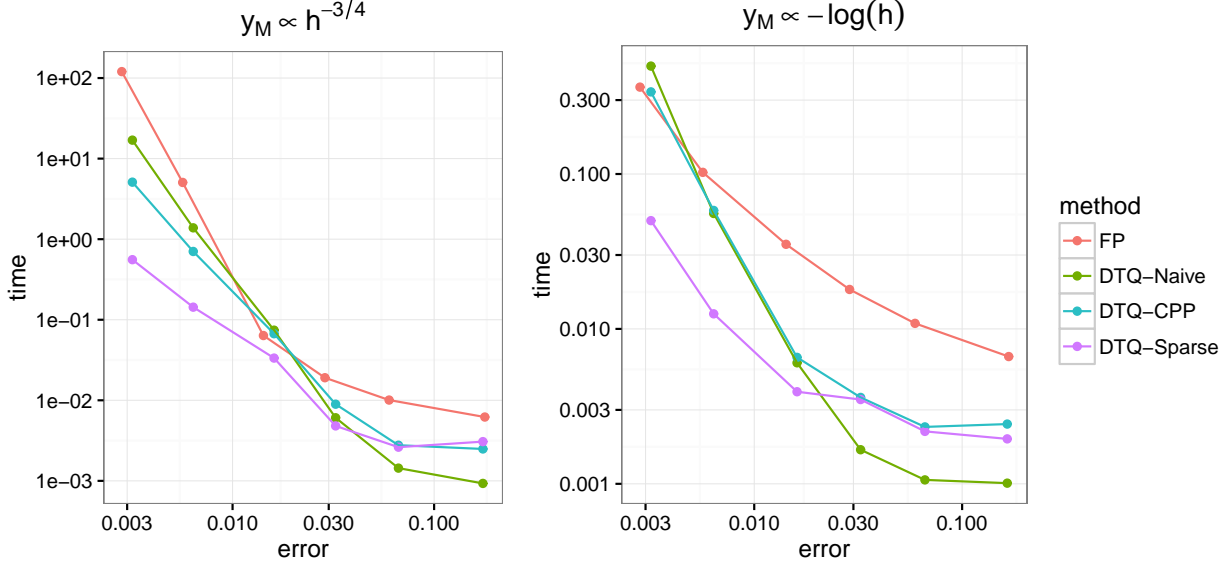


Figure 2: For a particular SDE, Example 1 from (64), suppose we are interested in computing the density  $p(x, T)$  at time  $T = 1$ . When we compute this density, we will incur some error, measured here in the  $L^1$  norm. **The plotted results show that for a fixed value of this error, the DTQ methods require less computational time (measured in wall clock seconds) than a method for numerically solving the Fokker-Planck PDE.** In all simulations, we use a domain  $[-y_M, y_M]$ . For the simulations in the left (respectively, right) plot, we have scaled the domain according to  $y_M \propto h^{-3/4}$  (respectively,  $y_M \propto \log h^{-1}$ ), where  $h > 0$  is the time step. In both plots, we see that for smaller values of the error, the fastest method is DTQ-Sparse; for larger values of the error, the fastest method is DTQ-Naïve. In particular, for the smallest error of 0.003, the DTQ-Sparse method is over  $10^2$  (respectively,  $10^{3/4}$ ) times faster than the Fokker-Planck method in the left (respectively, right) plot. Despite the fact that our Fokker-Planck solver uses the same sparse numerical linear algebra as DTQ-Sparse, it is often the slowest of the four methods. For details regarding the three implementations of the DTQ method (DTQ-Naïve, DTQ-CPP, and DTQ-Sparse) as well as the implementation of our Fokker-Planck solver, please see Section 7.2.